DOCUMENT RESUME

ED 224 804 TM 810 968

AUTHOR	Beaton, Albert E.
TITLE	Interpreting Least Squares without Sampling Assumptions.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-81-38
PUB DATE	Oct 81
NOTE	80p.
AVAILABLE FROM	Educational Testing Service, Research Publications R116, Princeton, NJ 08541.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Goodness of Fit; *Least Squares Statistics; *Mathematical Models; Measurement Techniques; *Regression (Statistics); Research Methodology; *Sampling; Statistical Analysis
IDENTIFIERS	Data Interpretation; *Inference (Statistical); *Linear Models; Randomization

ABSTRACT

Least squares fitting process as a method of data reduction is presented. The general strategy is to consider fitting (linear) models as partitioning data into a fit and residuals. The fit can be parsimoniously represented by a summary of the data. A fit is considered adequate if the residuals are small enough so that manipulating their signs and locations does not affect the summary more than a pre-specified amount. The effect of the residuals on the summary is shown to be (approximately) characterized by the output of standard regression programs. The general process of linear fitting models by least squares is covered in detail and discussed briefly in its relationship to standard hypothesis testing and to Fisher's randomization test. Fitting in weighted least squares and a comparison of fitting to standard methods are also discussed. It is shown that some of the output (e.g., standard errors, t, F, and p statistics) from standard regression programs can be interpreted as approximate measures of goodness-of-fit of a model to the observed data. The interpretation is also applicable in weighted least squares situations such as robust regression. (PN)

RR-81-38

RESEARCH REPORT

INTERPRETING LEAST SQUARES WITHOUT SAMPLING ASSUMPTIONS

Albert E. Beaton

October 1981

Educational Testing Service
Princeton, New Jersey

INTERPRETING LEAST SQUARES WITHOUT SAMPLING ASSUMPTIONS

Albert E. Beaton

Educational Testing Service
Princeton, New Jersey
October 1981

ABSTRACT

Least squares fitting is perhaps the most commonly used tool of statisticians. Under sampling assumptions, statistical inference makes possible the estimation of population parameters and their confidence intervals and also the testing of hypotheses. In this paper the properties of least squares fitting is examined without sampling assumptions. It is shown that some of the output (e.g. standard errors, t, F, and p statistics) from standard regression programs can be interpreted as (approximate) measures of goodness-of-fit of a model to the observed data. The interpretation is also applicable in weighted least squares situations such as robust regression.

ACKNOWLEDGMENTS

Albert E. Beaton is Director, Division of Measurement, Statistics, and Data Analysis Research, Educational Testing Service, Princeton, New Jersey 08540. The author is indebted to Thomas Kellaghan of the Educational Research Centre, Dublin, Ireland, for his support; to Paul Holland, Donald Rubin, and John Tukey for many helpful suggestions; particularly to Alfred Rogers and John Barone for help in programming; and Elaine Guennel and Ruth Myernick for typing.

1. INTRODUCTION

Apparently, fitting simple functions to observed data is a very basic human function. Art students are taught that an observer will see familiar shapes such as lines, squares, or circles even when only a few points or segments are actually present. Science students are taught that finding simple functions that represent complex events is one of the fundamental procedures of science. Given the breadth of application, it is no wonder that fitting functional forms to data has such an important position in statistical theory.

Statistical theory enhances fitting procedures in many ways. Given assumptions about a population and sampling procedures, we may infer that the fit from a random sample is an unbiassed estimator of a population parameter, and we can estimate the sample-to-sample variance of the parameter estimate. Under correct conditions, we may develop a confidence interval for a parameter or test an hypothesis that the parameter is some known value. Clearly, statistical theory adds substantially to the interpretation of fitting--at the cost of collecting (or assuming) a random sample and making assumptions about the population distribution of residuals. Statistical theory also dictates, to some degree, what functions of the data (e.g. standard error of a mean) that we interpret.

However, the estimation of population parameters is but one of the important functions of statistics. R. A. Fisher (1930) wrote on the first page of Statistical Methods of Research Workers, "Statistics may be regarded as (i) the study of populations, (ii) as the study of variations, (iii) as the study of the methods of the reduction of data." He also wrote of (page 5) "the practical need to reduce the bulk of any given body of data," and later (page 6) "We want to be able to express all the relevant information contained in the mass by means of comparatively few numerical values." Although random sampling may be important in estimating population parameters, there is no reason to forego data reduction in nonrandom samples as long as one is careful not to make the inferences that only random sampling allows.

The purpose of this paper is to discuss the fitting process as a method of data reduction. No assumptions about the sampling process nor population distribution will be made. Clearly, whenever the usual statistical assumptions are plausible, standard procedures of statistical inference should be used, but the concern here is with data for which the assumptions are quite inappropriate. The general strategy is to consider fitting (linear) models as partitioning data into a fit and residuals. The fit can be parsimoniously represented by a summary of the data. A fit is considered adequate if the residuals are small enough so that manipulating their signs and locations does not affect the summary more than a

pre-specified amount. The effect of the residuals on the summary is shown to be (approximately) characterized by the output of standard regression programs. Decision rules for accepting a fit will be proposed, and these rules will be shown to be equivalent in large samples to standard hypothesis tests.

That some of the statistics from a regression analysis can be interpreted as descriptive statistics is known (e.g. regression coefficients, squared multiple correlation), but this paper also shows possible interpretations of the covariance of regression coefficients, t , F , and "probability" statistics. Not much work seems to have been done in this area, but a paper by Freedman and Lane (1978) does approach the problem of nonstochastic significance testing and, although they differ in purpose and approach, come to similar conclusions where their work overlaps with what is covered here, and the bootstrap of Efron (1979) is in a similar spirit.

The next section in this paper will cover in detail the general process of linear fitting models by least squares and discuss briefly its relationship to standard hypothesis testing and to Fisher's randomization test.. The following section will discuss fitting in weighted least squares and compare fitting to standard methods. Most proofs are relegated to the appendix.

2. CASE I: ORDINARY LEAST SQUARES

Let us assume that we have a set of data and wish to fit a linear model using the least squares criterion. The data requirements and the notation used in this paper are shown in Table 1. All data elements are known, fixed, finite, real numbers. The matrix W is a diagonal matrix of weights which will be discussed later and can be assumed equal to an identity matrix here. Least squares fitting is a matter of algebra and, as long as $X'X$ has an inverse, an unique equation can be fitted. The computation of the least squares coefficients can be performed using a standard regression program.

Insert Table 1 about here

The output of most regression programs consists not only of the regression coefficients but also of a number of other statistics associated with regression analysis. Table 2 contains a fairly extensive list of regression statistics which might be included in computer output and a formula for each. That computer programs differ in internal algorithm or precision does not concern us here; we will assume that enough precision is kept so that rounding error can be ignored. The derivation and interpretation of these statistics using sampling theory are too well known to repeat here (see, for example, Graybill 1961, Draper and Smith 1966, Daniel and Wood 1971, Searle 1971, etc.)

Insert Table 2 about here

The interpretation of some of these statistics without sampling theory is also well known. Fisher (1930) showed that a regression coefficient may be considered as a weighted average of the response variable y where the weights are a function of a regressor variable. Tukey's catchers (Beaton and Tukey 1974) may be thought of as sets of weights to be applied to the response variable in order to form partial regression coefficients. Thus, any regression coefficient may be conceived of as a weighted average of the response variable. The standard error of estimate is a measure of how well the linear model fits the data, although division by the number of degrees of freedom comes from sampling considerations whereas division by the sample size would seem more appropriate for descriptive purposes. The squared multiple correlation may be interpreted as a relative measure of goodness of fit. However, some common regression outputs, the $cov(\underset{\sim}{b})$, t_j , $p(t_j)$, F , and $p(F)$, are not usually interpreted as descriptive statistics but from consideration of the inferred behavior of different random samples. We will show below that these statistics also permit interpretation as measures of goodness of fit without recourse to sampling theory.

The premise of this paper is that data reduction is a suffi-

cient reason for data analysis. We wish to reduce a mass of data into a few numbers which characterize or summarize an interesting relationship between the response vector $\underset{\sim}{y}$ and the regressor matrix X . We often do this by fitting a linear (in the parameters) model by least squares. Since the regression coefficients are weighted averages of the response values, they may be considered as summary statistics. A summary will in most cases lose some of the information in the original data in the sense that the response variable cannot be exactly reconstructed from X and the summary, $\underset{\sim}{b}$. To judge the adequacy of the summary, we will develop measures of how well the original data can be reconstructed from the summary and how sensitive the summary is to the information lost in the data reduction.

Fitting equations to data may be considered as a way of partitioning a set of observations $\underset{\sim}{y}$ into two parts, a fit, $\hat{\underset{\sim}{y}}$, and residuals, $\underset{\sim}{e}$; that is,

$$\underset{(\text{data})}{\underset{\sim}{y}} = \underset{(\text{fit})}{\hat{\underset{\sim}{y}}} + \underset{(\text{residuals})}{\underset{\sim}{e}} \tag{2.1}$$

where the values of $\hat{\underset{\sim}{y}}$ are related to the regressors X by the linear function $\hat{\underset{\sim}{y}} = X\underset{\sim}{b}$ and the residuals $\underset{\sim}{e} = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$ are what is left over. A data summary $\underset{\sim}{b}$ is considered good if $\hat{\underset{\sim}{y}} \cong \underset{\sim}{y}$, that is, the actual values $\underset{\sim}{y}$ are satisfactorily reconstructed from the fitted values which implies that $\underset{\sim}{e} \cong 0$ and thus may be ignored.

The residuals are, of course, minimal in the least squares sense, but minimal does not necessarily mean small.

We will here consider the vector $\tilde{e}$ to be small if its elements e_i are so close to zero that we can be indifferent to

1. changes in the signs[1] of the e_i , and
2. rearrangements of the locations of the e_i .

that is, we will consider the vector $\tilde{e}$ to be small if we can rearrange its elements and change some or all of their signs to form a vector $\tilde{e}_k$, say, then create a pseudo-data vector $\tilde{y}_k = \tilde{y} + \tilde{e}_k$ and still have the vector $\tilde{y} \cong \tilde{y}_k$. We will judge the closeness of $\tilde{y}_k$ to $\tilde{y}$ by summarizing $\tilde{y}_k$ in the same way that $\tilde{y}$ was summarized, that is, regress $\tilde{y}_k$ on X , and see whether the summary $\tilde{b}_k$, say, of $\tilde{y}_k$ is reasonably close to $\tilde{b}$. If a large proportion of all possible $\tilde{b}_k$ are close to $\tilde{b}$ in the sense discussed below, then we will consider the fit to be adequate.

There is a large number of ways in which the signs and locations of the e_i can be altered; there are 2^N different possible arrangements of the signs and $N!$ different ways to permute the elements of $\tilde{e}$, thus there are

$$K = 2^N N!$$

(not necessarily distinct) possible signed permutations of $\tilde{e}$.

Let us denote each possible signed permutation of $\underset{\sim}{e}$ as $\underset{\sim}{e}_k$ where $k = 1,2,\ldots,K$. The order in which the signed permutations are arranged is not important here, but for convenience we will denote $\underset{\sim}{e}_1 = \underset{\sim}{e}$, the vector with elements in the original order and with no sign changes.

It is convenient to write the $\underset{\sim}{e}_k$ as the product of a signed permutation matrix P_k and the vector $\underset{\sim}{e}$, i.e.

$$\underset{\sim}{e}_k = P_k \underset{\sim}{e} \tag{2.2}$$

where P_k is of order NxN , has one nonzero element in each row and column, and that nonzero element is +1 or -1. The location of the nonzero elements determine the permutation and the ± 1 determines sign changes. Since $\underset{\sim}{e}_1 = \underset{\sim}{e}$, P_1 is an NxN identity matrix.

In this notation, the pseudo-data vector

$$\underset{\sim}{y}_k = \underset{\sim}{y} + P_k \underset{\sim}{e} \tag{2.3}$$

and the summary computed from $\underset{\sim}{y}_k$ is

$$\underset{\sim}{b}_k = C' \underset{\sim}{y}_k = \underset{\sim}{b} + C' P_k \underset{\sim}{e} \quad . \tag{2.4}$$

The judgment of goodness of fit will be based on the differences between the $\underset{\sim}{b}_k$ and $\underset{\sim}{b}$.

The notation for ordinary least squares is summarized in Table 3.

Insert Table 3 about here

To illustrate the signed permutation scheme, a numerical demonstration is shown in Figures A and B. Figure A shows a simple data set with $N = 5$ and $m = 1$; thus, the regression line has two coefficients, an intercept and a slope. The coefficients of the best fit line, $\underset{\sim}{b}' = (2,3)$, as well as the fit $\hat{\underset{\sim}{y}}$, the residuals $\underset{\sim}{e}$, the catchers C , and the variance of the residuals σ^2 , are shown. If these residuals are small, then we should be able to resign and rearrange the elements of $\underset{\sim}{e}$ and then construct data sets $\underset{\sim}{y}_k$, and the summaries of these data sets, $\underset{\sim}{b}_k$, should be reasonably close to $\underset{\sim}{b}$. Figure B shows four such signed permutations. The vector $\underset{\sim}{e}_2$ is the same as $\underset{\sim}{e}$ except that the sign of the first element is changed; the pseudo-data vector $\underset{\sim}{y}_2 = \hat{\underset{\sim}{y}} + P_2\underset{\sim}{e}$ is also shown, as are the regression coefficients $\underset{\sim}{b}'_2 = (-9.2, 5.8)$ which result from the regression of $\underset{\sim}{y}_2$ on $\underset{\sim}{X}$. Comparing $\underset{\sim}{b}_2$ to $\underset{\sim}{b}$ shows that changing the sign of just one element of $\underset{\sim}{e}$ results in a regression line in which the intercept has a different sign. $\underset{\sim}{e}_3$ contains the same elements as $\underset{\sim}{e}$ except that the first two elements are exchanged; the sign of the intercept in the resulting regression coefficients $\underset{\sim}{b}_3$ is different from $\underset{\sim}{b}$. In the two other examples, P_3 was chosen so as to identify the signed permutation with the

largest value of the intercept and P_4 was chosen to maximize the slope.[2]

Insert Figures A and B about here

These are but four values of $\underset{\sim}{e}_k$; Figures C1 and C2 show the distribution of all $K = 3840$ different possible values of the slope and the intercept. The values of the intercept vary from -10 to +14 and about 38 percent differ in sign from the intercept computed from the unmodified data; the values of the slope range from -.6 to +6.6 with but about two percent differing in sign from the original slope. The residuals are large enough so that their signed permutation often results in intercepts which do not even have the same sign as the original, although the residuals are not so large as to affect the sign of the slopes in a vast majority of cases.

Insert Figures C1 and C2 about here

If we consider the coefficients of a fit adequate when the residuals are small enough such that the signed permutations of the residuals will seldom, if ever, result in regression coefficients with a different sign from the original, then the following

one-tailed decision rule for a single coefficient seems appropriate:

Decision Rule 1: The single regression coefficient b_j will be considered adequate if $100(1-\alpha)\%$ of the coefficients b_{kj} have the same sign as b_j, and inadequate otherwise, where α is a constant selected by the fitter.

Values such as .05 and .01 seem appropriate for α. If we are concerned with large deviations $b_{kj} - b_j$ in either direction, we may use a two-tailed decision rule:

Decision Rule 2: The fit of a regression coefficient b_j will be considered adequate if $100(1-\alpha)\%$ of the coefficients b_{kj} are no farther away in either direction from b_j than the point where the b_{kj} have a different sign from b_j, and inadequate otherwise, where α is a constant selected by the fitter.

More stringent fitting rules are possible; for example, we might require that the residuals be so small that their signed permutation does not affect the first (second, etc.) significant digit of a coefficient.

In addition to assessing the adequacy of fit for each individual regression coefficient, we may wish to judge the adequacy of fit of the vector $\underset{\sim}{b}$ taken as a whole. We might ask: in what proportion of the cases did the signed permutation scheme result in vectors $\underset{\sim}{b}_k$ in which all of the elements had different signs from $\underset{\sim}{b}$?

For the numerical example, the joint distribution of the slopes and intercepts is shown in Figure C3; the absence of any points in the third quadrant indicates that in no case did any $\underset{\sim}{b}_k$ differ from $\underset{\sim}{b}$ in all signs. However, a more interesting question might be: what proportion of cases are as far away <u>in any direction</u> as the origin, that is, the point beyond which all elements of $\underset{\sim}{b}_k$ have different signs. To answer this question, we need a definition of distance. Inspection of Figure C3 shows that the values of the slopes and intercepts are not independent; in fact they are correlated at about -.90. The choice of a particular $\underset{\sim}{e}_k$ affects the elements of $\underset{\sim}{b}_k$ in a complementary manner; it seems natural, therefore, to measure (squared) distance in a Mahalanobis-like manner, that is, the squared distance of any vector $\underset{\sim}{b}_k$ from $\underset{\sim}{b}$ is

$$d_k^2 = \left(\underset{\sim}{b}_k - \underset{\sim}{b}\right)' \left(\mathrm{cov}\left(\underset{\sim}{b}_k\right)\right)^{-1} \left(\underset{\sim}{b}_k - \underset{\sim}{b}\right) \qquad (2.5)$$

and, in the same metric, the squared distance of $\underset{\sim}{b}$ from the point beyond which all elements have different sign is

$$d^2 = \underset{\sim}{b}' \left(\mathrm{cov}\left(\underset{\sim}{b}_k\right)\right)^{-1} \underset{\sim}{b} \qquad (2.6)$$

The distribution of all possible values of d_k^2 is shown in Figure C4.

Insert Figures C3 and C4 about here

Using this definition, we may form the following decision rule:

Decision Rule 3: The fit of the regression coefficients $\underset{\sim}{b}$ will be considered adequate if $100(1 - \alpha)\%$ of the vectors $\underset{\sim}{b}_k$ are no farther away from $\underset{\sim}{b}$ than the point where all elements of $\underset{\sim}{b}_k$ have different signs from those of $\underset{\sim}{b}$, and inadequate otherwise, where α is a constant selected by the fitter.

It is unusual in statistical applications to be concerned about the fit of all coefficients; in practice, the intercept is often arbitrary and not of interest. Decision Rule 3 can be modified so that it refers to any subset of the regression coefficients.

Using these decision rules directly implies calculating all K possible values of $\underset{\sim}{b}_k$ and this is clearly impossible except in very small samples; in fact, for a modestly sized sample in which $N = 30$, the number of possible sets of regression coefficients is $K \cong 2.8 \times 10^{41}$ and thus cannot be calculated by simple direct measures. Conceivably, Monte Carlo techniques could be used to estimate the proportion of $\underset{\sim}{b}_k$ in any particular range, but such would entail a large investment in computer time and programming. We can, however, approximate the decision statistics for reasonably large samples.

Let us first examine the distribution of the $\underset{\sim}{b}_k$. We can calculate the mean and covariance of $\underset{\sim}{b}_k$ since

$$\text{ave}\left(\underset{\sim}{b}_k\right) = \underset{\sim}{b} \tag{2.7}$$

$$\text{cov}\left(\underset{\sim}{b}_k\right) = \sigma^2 (X'X)^{-1} \tag{2.8}$$

and the skewness and kurtosis of the jth element of $\underset{\sim}{b}_k$ are

$$\text{skew}\left(b_{kj}\right) = 0 \tag{2.9}$$

$$\text{kurt}\left(b_{kj}\right) = \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right) \tag{2.10}$$

where β_{2e} and β_{2c_j} are the measures of the kurtosis of $\underset{\sim}{e}$ and the jth column of the catcher matrix respectively. (See Table 9 for summary of definitions.) Note that these calculations are exact, not estimates or approximations. The skewness is exactly that of a normal distribution (as are all odd moments), and it is easily seen that, given fixed values of the kurtosis of the residuals and catcher vectors, that

$$\lim_{N \to \infty} \text{kurt}\left(b_{kj}\right) \to 3 \tag{2.11}$$

that is, as the sample size grows large, the kurtosis of each b_{kj} approaches the kurtosis of the normal distribution. Thus, the distribution of the b_{kj} has, in the limit, the same characteristics as the normal curve with mean b_j and variance $\sigma_{b_j}^{\ 2} = \sigma^2 (X'X)^{jj}$ where $(X'X)^{jj}$ is the jth diagonal element of $(X'X)^{-1}$.

If we accept the normal distribution as close enough to the distribution of b_{kj} for our purposes, then we can use a table of the normal curve to approximate the proportion of b_{kj}

within any particular range. Let us assume that we wish to approximate the proportion of b_{kj} that differ in sign from b_j which was computed in the orginal solution. Since we know the mean and variance of the b_{kj} exactly, we can form the standard normal deviate

$$z_j = \frac{b_j}{\sqrt{\sigma^2 (X'X)^{jj}}}$$

which can be referred to a table of the normal curve to find the approximate proportion $\tilde{p}\left(z_j\right)$. If $\tilde{p}\left(z_j\right)$ is one-tailed, then it is approximately the proportion of b_{kj} with different sign than b_j . If $\tilde{p}\left(z_j\right)$ is two-tailed, then it is approximately the proportion of b_{kj} as far away from b_j as the point beyond which the b_{kj} differ in sign from b_j . The values of $\tilde{p}\left(z_j\right)$, therefore, can be used as approximations for the values needed for Decision Rules 1 and 2.

We can also develop an approximation for the proportion of $\underset{\sim}{b}_k$ as far away from $\underset{\sim}{b}$ as the point where all the elements of $\underset{\sim}{b}_k$ differ in sign from $\underset{\sim}{b}$. The mean and variance of the squared distances of the b_k from $\underset{\sim}{b}$ are

$$\text{ave}\left(d_k^{\,2}\right) = m + 1 \tag{2.12}$$

$$\text{var}\left(d_k^{\,2}\right) = \left(\frac{1 - \beta_{2e}}{N - 1}\right)(m + 1)^2 + 2\left(\frac{N - \beta_{2e}}{N - 1}\right)(m + 1) + \Sigma_i q_{ii}^{\,2}\left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1}\right) \tag{2.13}$$

The proofs are in the appendix. Both the mean and variance are exact, not estimates or approximations. The mean is exactly that of the χ^2 distribution with $m + 1$ d.f. Given a fixed kurtosis of the residuals, β_{2e} , then

$$\lim_{N \to \infty} \operatorname{var}\left(d_k^{\,2}\right) \to 2(m + 1) \tag{2.14}$$

since $\Sigma_i q_{ii}^{\,2} \to 0$ as the sample size increases, thus the variance of the $d_k^{\,2}$ approaches the variance of the χ^2 distribution with $m + 1$ d.f. If we accept the chi-squared distribution as close enough to the distribution of the $d_k^{\,2}$ for our purposes, then the squared distance d^2 of the point where all elements of $\underset{\sim}{b}_k$ have different sign from $\underset{\sim}{b}$ can be referred to a table of the chi-squared distribution for an approximation of the percent of $\underset{\sim}{b}_k$ as far away as the origin or, alternatively, the statistic

$$F^* = \frac{d^2}{\text{ave } d_k^{\,2}} = \frac{b'\left(\operatorname{cov}\left(b_k\right)\right)^{-1} b}{m + 1} = \frac{y'x(X'X)^{-1}x'y}{\sigma^2(m + 1)} \tag{2.15}$$

may be computed and referred to an F table with $m + 1$ and ∞ d.f. Let us call this proportion $\tilde{p}(F^*)$. $\tilde{p}(F^*)$ is thus an approximation of the value for Decision Rule 3.

As mentioned above, in many statistical applications it will be of interest to measure the goodness-of-fit of some subset of $\underset{\sim}{b}$ instead of the entire vector; for instance, a researcher may be interested in the goodness-of-fit of the slopes in a multiple regression but not in the fit of the intercept. Let us call the subset of

interest $\underset{\sim}{b}_s$ which is of length m_s $\left(m_s \leq m + 1\right)$. Let us call $\underset{\sim}{b}_{ks}$ the equivalent subset of $\underset{\sim}{b}_k$, the kth signed permutation summary. The question of interest is: what proportion of the $\underset{\sim}{b}_{ks}$ are as far away from $\underset{\sim}{b}_s$ as the point where all signs in $\underset{\sim}{b}_{ks}$ are different from $\underset{\sim}{b}_s$? The squared distance of the vectors $\underset{\sim}{b}_{ks}$ from $\underset{\sim}{b}_s$ is

$$d_{ks}^{\;2} = \left(\underset{\sim}{b}_{ks} - \underset{\sim}{b}_s\right)' \left(\mathrm{cov}\left(\underset{\sim}{b}_{ks}\right)\right)^{-1}\left(\underset{\sim}{b}_{ks} - \underset{\sim}{b}_s\right) \tag{2.16}$$

and the distance of the point where all $\underset{\sim}{b}_{ks}$ elements change sign, the origin, from $\underset{\sim}{b}_s$ is

$$d_s^{\;2} = \underset{\sim}{b}_s'\left(\mathrm{cov}\left(\underset{\sim}{b}_{ks}\right)\right)^{-1} \underset{\sim}{b}_s \tag{2.17}$$

Corollary 1 in the appendix shows that the mean and variance of $d_{ks}^{\;2}$ are

$$\mathrm{ave}\left(d_{ks}^{\;2}\right) = m_s \tag{2.18}$$

$$\mathrm{var}\left(d_{ks}^{\;2}\right) = \left(\frac{1 - \beta_{2e}}{N - 1}\right) m_s + 2\left(\frac{N - \beta_{2e}}{N - 1}\right) m_s + \Sigma_i q_{ii}^{\;2}\left(\frac{\beta_{2e}(N + 1)}{N - 1} - \frac{3N}{N - 1}\right) \tag{2.19}$$

The mean is the same as the χ^2 distribution with m_s d.f. and the variance in its limit

$$\lim_{N \to \infty} \mathrm{var}\left(d_{ks}^{\;2}\right) \to 2m_s \tag{2.20}$$

approaches the variance of the chi-squared distribution, thus, if we accept the chi-squared distribution for approximation purposes, we may look up d^2 in a chi-square table or, alternately, compute

$$F^* = \frac{d_s^{\,2}}{\mathrm{ave}\left(d_{ks}^{\,2}\right)} = \frac{\underset{\sim}{b}_s\left(\mathrm{cov}\left(d_{ks}^{\,2}\right)\right)^{-1} \underset{\sim}{b}_s}{m_s} \tag{2.21}$$

which may be referred to the F table with m_s and ∞ d.f. The resultant value, $p(F^*)$ is approximately the proportion of $\underset{\sim}{b}_{ks}$ as far or farther away from $\underset{\sim}{b}_s$ as the origin.

If the researcher is interested in how close the values of the fits $\hat{\underset{\sim}{y}}_k$ are to the original data points $\underset{\sim}{y}$, we can show that the

$$\mathrm{ave}\left(\hat{\underset{\sim}{y}}_k\right) = \hat{y} \tag{2.22}$$

$$\mathrm{cov}\left(\hat{\underset{\sim}{y}}_k\right) = \sigma^2 X(X'X)^{-1} X' \tag{2.23}$$

Again the normal distribution, the proportion of y_{ki} in any interval can be approximated.

Table 4 summarizes the approximation scheme for the numerical example. It should be remembered that the sample size is very small, five, and that the residuals e_i are not close to normally distributed; the summary here is to show numerical calculations, not the adequacy of the approximations. This table first shows the mean, variance, skewness, and kurtosis for the distribution of

b_{k0}, the intercept, and b_{k1}, the slope, and the mean and variance of the d_k^2, the squared distances. The statistics for the approximation of the proportion of b_{kj} with different signs from the b_j are

$$z_0 = 2/\sqrt{29.48} = .368$$

$$z_1 = 3/\sqrt{2.68} = 1.833$$

which result in the (one-tailed) proportions $\tilde{p}(z_0) = .37$ and $\tilde{p}(z_1) = .03$ which are reasonably close to the actual proportions $p(b_0)$ and $p(b_1)$ which were computed by counting. The value of F^*

$$F^* = 25.9328/2 = 12.966$$

should be referred to an F table with two and ∞ d.f. from which the value $p(F^*)$ is found to be close to zero. The actual proportion is exactly zero. Thus, according to the decision rules, we would consider the intercept inadequate and consider the slope adequate or inadequate depending on whether we used Decision Rule 1 or 2. The equation as a whole would be considered to fit adequately.[3]

Insert Table 4 about here

2.1 Comment

It is interesting to note the similarities and differences between classical hypothesis testing and the signed permutation approach to goodness of fit. Classical hypothesis testing usually assumes that there is a model in some population of the form $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, that the elements ε_i of $\underset{\sim}{\varepsilon}$ are i.i.d. $N\left(0,\sigma_p^2\right)$ where σ_p^2 is the variance of the ε_i in the population, and that the sample at hand was randomly selected from that population; in return for these assumptions, probabilistic statements about the unknown values of $\underset{\sim}{\beta}$ can be made. The signed permutation approach makes no assumptions about the world outside of the sample and thereby forfeits the right to make statements outside of the sample itself. Given these major differences, it is interesting to note that these two approaches lead to very nearly the same calculations. Under its assumptions, the usual hypothesis $\underset{\sim}{\beta}_s = 0$ (where β_s includes all β_j except the intercept) leads to the statistics shown in Table 2 and the only differences between these and signed permutation statistics are

$$\sigma^2 = \frac{N - m - 1}{N} s_e^2 \tag{2.24}$$

$$j = \sqrt{\frac{N}{N - m - 1}}\, t_j \tag{2.25}$$

$$F^* = \frac{N}{N - m - 1} F \tag{2.26}$$

$$\mathrm{cov}\left(\hat{\underset{\sim}{b}}_k\right) = \frac{N - m - 1}{N} \mathrm{cov}\left(\underset{\sim}{b}\right) \tag{2.27}$$

and referring t_j to a Student's t table and F to an F table with N-m-1 d.f. in the denominator. For large N and moderate m , the factor N/(N-m-1) is trivial as is the difference between the t and normal curve tables and the difference in degrees of freedom in the denominator for the F table. Thus, the two approaches lead to the same probability statistics.

This leads to the interesting fact that the probability of finding a value as large (small) or larger (smaller) than the sample value b_j if the value of $\beta_j = 0$ in the population is the same as the proportion of signed permutations b_{kj} with different signs than b_j . Also, the probability of finding a vector $\underset{\sim}{\beta}_s$ as far away as $\underset{\sim}{b}_s$ by chance if the equivalent subset of $\underset{\sim}{\beta}_s$ in the population was zero is the same in the limit as the number of $\underset{\sim}{b}_{ks}$ as far away from $\underset{\sim}{b}_s$ as the point where all elements of $\underset{\sim}{b}_{ks}$ have different signs from $\underset{\sim}{b}_s$.

That the two approaches arrive at similar places should not be surprising. The usual sampling assumptions that $E(\underset{\sim}{\varepsilon}) = 0$ and $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma_p^2 I$ are analogous to lemmas 5a and 5b which are that $K^{-1}\Sigma_k \underset{\sim}{a}_k = 0$ and $K^{-1}\Sigma_k a_k a'_1 = \sigma^2 I$ since the average residual is zero in ordinary least squares. With the substitution of the expectations into the theorems in the appendix, the same theorems would almost suffice for hypothesis testing. The major difference would be the $E(s_e^2)$ which would result in the need for a correction for degrees of freedom and this correction would be

carried to the $\text{cov}(\hat{\underset{\sim}{b}})$ and statistics derived from it.

From the computational point of view, the statistics from a standard regression program are close enough to the statistics for signed permutations that they can be used directly in interpreting goodness of fit if the sample size is much larger than the number of variables.

2.2 Randomization Tests

The Fisher (1926)-Pitman (1937) method of randomization has great appeal because it is derived solely from the mechanics of randomization without any assumptions about the parent population. Basically, Fisher permutes residuals about a null model in which all parameters are specified and, since the parameters are often specified to be zero, the residuals may be the original data values $\underset{\sim}{y}$. Since sampling is involved, there is a major difference in approach between a randomization test and measuring goodness of fit by signed permutations.

It is possible to view the signed permutation scheme as a variation of a randomization test. First, the population from which the sample was selected would have to be assumed to be symmetric to justify assuming that $-e_i$ was as likely to occur as e_i . Secondly, the residuals about the null model would be resigned and permuted instead of the residuals about the completely fitted model as done when measuring goodness of fit. All of the parameters would have to be specified since if any parameters were

fit from the data the residuals would be overfitted.[4] With these adaptations, a distribution of all possible results of an experiment could be generated (for small N) and the probability of any particular result calculated. The moments of the distribution of results are easily calculated using the lemmas in the appendix.

3. WEIGHTED LEAST SQUARES

Weighted least squares is another common statistical technique which has been used for a number of different purposes. A few uses are:

1. Equalizing variance. In samples where the variances of residuals are known a priori to be different at different parts of the regression surface, the variances can be equalized using weights which are proportional to the inverse of the square root of the variance. In this case, the sampling assumptions and inferences are well known.

2. Differential precision. In some fittings, the researcher may be interested in fitting some segment of the regression plane more carefully than other segments and does so by weighting the residuals in that segment more heavily than the rest. This is similar to the situation in survey research where various strata are sampled with different probabilities to assure a fixed representation of all strata, and the inverse of the sampling probability is used as a weight when combining strata.

3. Robust/resistant regression. Robust/resistant regression uses the sample residuals to form weights and, by discounting large residuals, reduces the effect of outliers on the fit (see Mosteller and Tukey 1977; Beaton and Tukey 1974). Robust/resistant regression may modify weights iteratively.

The implications from sampling theory are quite different in the above cases, but the basic fitting procedure is the same. We will again use the definitions of data in Table 1. Note that the weights W are considered fixed as well as the data X and $\underset{\sim}{y}$. Where the weights came from is not important here; it is immaterial whether W came from sampling considerations or from iteratively reweighting residuals. All that is required is that $X'WX$ has an inverse.

The basic algebra of the fitting process is clear. If we wish to fit a linear model of the form

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e} \tag{3.1}$$

subject to the condition that $e'We$ be a minimum, then the value of $\underset{\sim}{b}$ which minimizes $e'We$ is

$$\underset{\sim}{b} = (X'WX)^{-1}X'Wy \quad . \tag{3.2}$$

Thus the value of the fit is

$$\hat{y} = X\underset{\sim}{b} \tag{3.3}$$

and the residuals are

$$\underset{\sim}{e} = \underset{\sim}{y} - \hat{\underset{\sim}{y}} = \underset{\sim}{y} - X\underset{\sim}{b} \qquad (3.4)$$

Weighted least squares, then may also be considered as partitioning $\underset{\sim}{y}$ into two parts

$$\underset{(\text{data})}{\underset{\sim}{y}} = \underset{(\text{fit})}{\hat{\underset{\sim}{y}}} + \underset{(\text{residuals})}{\underset{\sim}{e}} \qquad (3.5)$$

If the fit $\hat{\underset{\sim}{y}}$ and the summary $\underset{\sim}{b}$ are to be accepted as adequate, then we should assure ourselves that the residuals are not large enough to affect seriously either. As with ordinary least squares, we may develop a metric of goodness of fit by examining the effect of signed permutations of the residuals.

However, there are at least two reasonable ways in which the residuals may be resigned and permuted:

Case II: Weight the resigned and permuted residuals, and

Case III: Resign and permute the weighted residuals.

Both of these cases use the same data summary, $\underset{\sim}{b}$, which was computed so as to minimize the objective function $\underset{\sim}{e}'W\underset{\sim}{e}$, but differ in the way that the pseudo-data sets are constructed.

In Case II, the resigning and permuting is done without any reference to the weights. The weights are considered to be associated with the rows $\underset{\sim}{x}_i$ of X , thus, whatever residual is attached to the $\hat{\underset{\sim}{y}}_i$ computed from $\underset{\sim}{x}_i\underset{\sim}{b}$ will be weighted by

the weight associated with the ith observation. This approach seems reasonable when the weights are used for differential precision since whatever residuals associated with a sensitive area will be so weighted.

In Case III, the weights are associated with the residuals, so that the weighting is done before the resigning and permuting. The weighted residuals are therefore added to value of $\underset{\sim}{x}_i b$ to compute the pseudo-data sets. This approach seems more appropriate when the weights are chosen to operate on the residuals as when chosen to equalize the variance of residuals or in robust/ resistant regression.

3.1 Case II: Weighting the Resigned and Permuted Residuals

The basic definitions for Case II are shown in Table 5. Although these definitions are quite similar to the definitions for ordinary least squares (Case I), there are some important differences which should be noted. The weights are included in the catchers, i.e. $C = WX(X'WX)^{-1}$. The mean of the residuals is not in general equal to zero, although the weighted mean, that is, the mean of $W\underset{\sim}{e}$, is. The value μ_{2e} does not equal the variance of the residuals but is simply the unweighted, uncentered mean square of the residuals.

Insert Table 5 about here

The pseudo-values of the vector y are formed by

$$\underset{\sim}{y}_k = \hat{\underset{\sim}{y}} + P_k\underset{\sim}{e} \tag{3.6}$$

that is, a signed permutation of the residuals is added to the fit as in Case I, and the data are summarized by

$$\underset{\sim}{b}_k = C'\underset{\sim}{y}_k = \underset{\sim}{b} + C'P_k\underset{\sim}{e} \tag{3.7}$$

The question is still whether or not the elements of $\underset{\sim}{b}_k$ often differ in sign from the corresponding elements of $\underset{\sim}{b}$.

Figure D contains the results of a numerical demonstration of Case II. The weights were arbitrarily chosen. The weighted regression coefficients $\underset{\sim}{b}$ and the other basic statistics of Case II are shown in the top of the figure. Note that the intercept using these weights is negative whereas the unweighted intercept is positive. Four signed permutations are shown. P_2 and P_3 are the same signed permutations as used in Case I, but only one results in a difference in sign from the original summary. P_4 and P_5 were chosen to display the largest intercept and slope respectively.

Insert Figure D about here

Since the number of signed permutations is the same in weighted regression, we cannot reasonably compute all $\underset{\sim}{b}_k$ and count the

number of b_{kj} with different signs from the b_j nor can we compute the number of d_k^2 as far away as the point where all elements have different signs, that is

$$d^2 = \underset{\sim}{b}'\left(\operatorname{cov}\left(\underset{\sim}{b}_k\right)\right)^{-1} \underset{\sim}{b} \tag{3.8}$$

However, we do know something about the distribution of these statistics. We know

$$\operatorname{ave}\left(b_k\right) = \hat{\underset{\sim}{b}} \tag{3.9}$$

$$\operatorname{cov}\left(b_k\right) = \mu_{2e} WX(X'WX)^{-1} X'W^2X(X'WX)^{-1} X'W \tag{3.10}$$

$$\operatorname{skew}\left(b_{kj}\right) = 0 \tag{3.11}$$

$$\operatorname{kurt}\left(b_{kj}\right) = \frac{\beta_{2e}\beta_{2c_i}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2e_j}}{N}\right) \tag{3.12}$$

$$\operatorname{ave}\left(d_k^{\,2}\right) = m + 1 \tag{3.13}$$

$$\operatorname{var}\left(d_k^{\,2}\right) = \left(\frac{1 - \beta_{2e}}{N - 1}\right)(m + 1)^2 + 2\left(\frac{N - \beta_{2e}}{N}\right)(m + 1) + \Sigma_i q_{ii}^{\,2}\left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1}\right) \tag{3.14}$$

$$\operatorname{ave}\left(\hat{y}_k\right) = \hat{\underset{\sim}{y}} \tag{3.15}$$

$$\operatorname{cov}\left(\hat{y}_k\right) = \mu_{2e}\ X(X'WX)^{-1} X'W^2X(X'WX)^{-1} X' \tag{3.16}$$

The proof is shown in Theorem 2 in the appendix.

It is easily shown that

$$\lim_{N \to \infty} \text{kurt}\left(b_{kj}\right) \to 3 \tag{3.17}$$

and thus as the sample grows large, given fixed β_{2e} and β_{2c_j}, the moments of the distribution of b_k approach the moments of the normal distribution. Also, the variance of d_k^2

$$\text{var}\left(d_k^{\,2}\right) \to 2(m + 1) \tag{3.18}$$

approaches the variance of the χ^2 distribution with $2(m + 1)$ d.f. If we accept the normal and χ^2 distributions as close enough for our purposes, then we can compute

$$z_j = b_j/\sigma_{b_{jj}}^{\;2} \tag{3.19}$$

and

$$F^* = d^2/(m + 1) \tag{3.20}$$

which can be referred to normal curve or F tables for approximate proportions to be used in the decision rules. The numerical results for Case II are shown in Table 6.

Insert Table 6 about here

3.2 Comment

Case II is analogous to the situation in sampling in which one makes the usual assumptions of ordinary least squares but

weights the data anyway. That is, one assumes that $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{e}$ where $\underset{\sim}{\beta}$ is the population parameters, and that $E(e) = 0$ and $E(ee') = \sigma_p^{\ 2}I$ where $\sigma_p^{\ 2}$ is the variance of $\underset{\sim}{e}$ in the population. Under these assumptions,

$$E\left(\mu_{2e}\right) = \sigma_p^{\ 2} \tag{3.21}$$

The value $\hat{b}$ is ab unbiassed estimator

$$E(\hat{b}) = \underset{\sim}{\beta} \tag{3.22}$$

and the sample to sample variance of b is

$$\text{cov}(\hat{b}) = \sigma_p^{\ 2}WX(X'WX)^{-1}\ X'W^2X(X'WX)^{-1}\ X'W \tag{3.23}$$

Thus the difference between the signed permutation approach and weighted regression using these assumptions results in using degrees of freedom as a divisor instead of the sample size N .

It is worth noting that Case II results in more complex and unusual calculations than Case I and, as we shall see, Case III. To compute the $\text{cov}\left(\underset{\sim}{b_k}\right)$ in one pass over the data requires some unweighted summations, summations using the weights, and summations using the squares of the weights. We know of no computer programs that come close to these calculations. Weighting for differential precision will, therefore, come at considerable cost.

3.3 Case III: Resigning and Permuting the Weighted Residuals

To examine the effect of the residuals in Case III, it is

convenient to transform the data in such a way that the data can be handled as if they were unweighted. This can be done by defining a diagonal matrix V in which the diagonal elements are the square roots of the equivalent elements in W, thus, $V^2 = V'V = W$. Using an asterisk to identify transformed data, we can form $\underset{\sim}{y}^* = V\underset{\sim}{y}$ and $X^* = VX$. The catchers are $C = X^*(X^{*\prime}X^*)^{-1} = VX(X'WX)^{-1}$ and the regression coefficients are

$$\underset{\sim}{b} = C'\underset{\sim}{y}^* = \underset{\sim}{b} \tag{3.24}$$

which is to say that the regression coefficients are not changed by the transformation. However, the fit

$$\underset{\sim}{\hat{y}}^* = X^*\underset{\sim}{b} = V\underset{\sim}{\hat{y}} \tag{3.25}$$

and the residuals

$$\underset{\sim}{e}^* = \underset{\sim}{y}^* - \underset{\sim}{\hat{y}}^* = V\underset{\sim}{e} \tag{3.26}$$

are. Note that the residuals e_i^* do not sum to zero and that $\mu_{2e} = \underset{\sim}{e}^{*\prime}\underset{\sim}{e}^* = \underset{\sim}{e}'W\underset{\sim}{e}$ which is the objective function which was minimized. The notation for Case III is summarized in both weighted and unweighted form in Table 7.

Insert Table 7 about here

Using the weighted data, the pseudo-values of y can be defined as

$$\underset{\sim}{y}_k^* = \hat{\underset{\sim}{y}}^* + P_k\underset{\sim}{e}^* \tag{3.27}$$

and the regression coefficients computed using these values are

$$\underset{\sim}{b}_k = C'\underset{\sim}{y}_k^* = \underset{\sim}{b} + C'P_k\underset{\sim}{e}^* \tag{3.28}$$

Figure E contains a numerical demonstration of this signed permutation scheme. The weights are the same as in Case II; thus the regression coefficients are the same. The values of y^*, X^*, and C are shown. The first two signed permutation matrices, P_2 and P_3, are the same as used in the previous cases and the last two, P_4 and P_5, were selected to result in the maximum intercept and slope respectively under this signed permutation scheme. The demonstration shows that the distribution of $\underset{\sim}{b}_k$ is quite different in Case III than Case II.

Insert Figure E about here

As in the previous cases, we cannot compute all K values of $\underset{\sim}{b}_k$ except in very small samples and thus we cannot calculate the proportions needed for using the decision rules. However, we again know the moments of the distribution exactly if we compute

$$\text{ave}\left(b_k\right) = b \tag{3.29}$$

$$\text{cov}\left(b_k\right) = \frac{e'we}{N}(X'WX)^{-1} \tag{3.30}$$

$$\text{skew}\left(b_{kj}\right) = 0 \tag{3.31}$$

$$\text{kurt}\left(b_{kj}\right) = \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right) \tag{3.32}$$

$$\text{ave}\left(d_k^{\;2}\right) = m + 1 \tag{3.33}$$

$$\text{var}\left(d_k^{\;2}\right) = \frac{1 - \beta_{2e}}{N-1}(m+1)^2 + 2\left(\frac{N - \beta_{2e}}{N-1}\right)(m+1) + \Sigma_i q_{ii}^{\;2}\left(\frac{\beta_{2e}(N+2)}{N-1} - \frac{3N}{N-1}\right) \tag{3.34}$$

$$\text{ave}\left(\hat{y}_k\right) = \hat{y} \tag{3.35}$$

$$\text{cov}\left(\hat{y}_k\right) = \frac{e'we}{N} X(X'WX)^{-1} X' \tag{3.36}$$

Proof is shown in Theorem 3. Since the $\text{kurt}\left(b_{kj}\right)$ and the $\text{var}\left(d_k^{\;2}\right)$ approach the same limits as before, we can compute

$$z_j = b_j \Big/ \sqrt{\text{diag}\left(\text{cov}\left(b_k\right)\right)} \tag{3.37}$$

and

$$F^* = d^2/(m+1) \tag{3.38}$$

which can be referred to normal curve or F tables for the approximate proportions to be used in the decision rules.

The numerical values for this signed permutation scheme are summarized in Table 8.

Insert Table 8 about here

3.4 Comment

Case III is analogous to the sampling situation in which one assumes that $y = X\beta + \varepsilon$ and that the $E(\varepsilon) = 0$ and that $E(\varepsilon\varepsilon') = \sigma_p^2 W^{-1}$ where W^{-1} is the inverse of the fixed matrix W . Under these assumptions,

$$E\left(\mu_{2e}\right) = \sigma_\rho^2 \tag{3.39}$$

The proofs that $E(b) = \beta$ and that the $\text{var}(\underset{\sim}{b}) = \sigma_p^2(X'WX)^{-1}$ are in Draper and Smith (1966).

It is worth noting that the computations necessary to use Case III are simple. The necessary summaries are $y'Wy$, $y'WX$, and $X'WX$ from which decision statistics can be computed. In Case III, as in Case II, the decision statistics are not affected by multiplying W by any positive constant.

Insert Table 9 about here

---

4. CONCLUSION

The signed permutations of residuals leads to an interpretation of least squares calculations without sampling assumptions but at the cost of the inability to generalize formally to a population Since in many applications the assumption of a random sample from a known distribution may be unwarranted or an attempt at an actual random sample may be thwarted by practical concerns, the interpretation of regression statistics as measures of fit in the obtained data may be the best that a researcher can do. Such an interpretation is not trivial, however, since substantial confidence in a model may come from fitting the model to many data sets under many conditions and finding that the resultant coefficients are reasonably similar.

Signed permutations may be most useful in robust/resistant regression where the iterative reweighting clearly violates the assumption of known, fixed weights and thus the sampling theory of weighted regression. At least the "probability statistics" generated from iterative reweighting have an interpretation as measures of goodness-of-fit. These goodness-of-fit measures may, therefore, be used as a criterion for the effectiveness of weighting systems or for comparing different systems now in use.

Footnotes

[1] The author is indebted to Paul Holland for suggesting the sign changes.

[2] The algorithm for finding the maximum values is due to Lustig (1979). To find the maximum b_{kj} , the column $\underset{\sim}{c}_j$ and the residuals are each rearranged into descending order of absolute magnitude and the signs of the residuals are changed to be the same as the corresponding elements of $\underset{\sim}{c}_j$.

[3] The analogous test in sampling theory is seldom used, that is, we seldom test the hypothesis that all parameters including the intercept are simultaneously equal to zero. The test of the subset containing only b_1 would result in an $F^* = z_1^2$.

[4] The sum of squared residuals is $\underset{\sim}{e}'\underset{\sim}{e} = \underset{\sim}{\varepsilon}'(I - X(X'X)^{-1}X'\underset{\sim}{\varepsilon}$ where $\underset{\sim}{\varepsilon}$ is the population residuals. Tukey has suggested weighting the e_i by the corresponding inverse square root of the diagonal of $(I - X(X'X)^{-1}X')$. This idea has not been followed up at this time.

1. Notation for Data

Statistic	Definition
N	number of observations
m	number of regressors
$i,i' = 1,2,\ldots,N$	indices of observations
$j,j' = 0,1,2,\ldots,m$	indices of regressors
$\underset{\sim}{y} = \{y_i\}$	Nth order vector of response values to be fitted
$X = \{x_{ij}\}$	N X (m + 1) matrix of values of regressors. All elements $x_{i0} = 1$. The rank of X is m + 1
$W = \{W_{ii'}\}$	NxN positive definite diagonal matrix of weights

NOTE: All data elements are fixed, known finite, real numbers.

2. Ordinary Least Squares Definitions

Statistic	Description
$\bar{y} = \Sigma_i y_i/N$	mean value of the y_i
$s_y^2 = \Sigma(y_i-\bar{y})^2/(N-1)$	variance of the y_i
$\hat{\underset{\sim}{b}} = \{b_j\} = (X'X)^{-1}X'\underset{\sim}{y}$	m+1 order vector of regression coefficients
$\hat{\underset{\sim}{y}} = \{\hat{y}_i\} = X\hat{\underset{\sim}{b}}$	Nth order vector of fitted values
$\underset{\sim}{e} = \{e_i\} = \underset{\sim}{y} - X\hat{\underset{\sim}{b}}$	Nth order vector of residuals
$s_e = \sqrt{\underset{\sim}{e}'\underset{\sim}{e}/(N-m-1)}$	standard error of estimate
$R^2 = \dfrac{\Sigma(\hat{y}_i-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$	squared multiple correlation
$F = \dfrac{\Sigma(\hat{y}_i-\bar{y})^2/m}{\Sigma(y_i-\hat{y}_i)^2/(N-m-1)}$	test statistic for $\beta_1=\beta_2=\ldots=\beta_m=0$
$p(F)$	probability associated with F
$\mathrm{cov}(\hat{\underset{\sim}{b}}) = \{\mathrm{cov}(\hat{b}_{jj'})\} = s_e^2(X'X)^{-1}$	m+1 by m+1 matrix of variances and covariances of $\hat{\underset{\sim}{b}}$
$SE(\hat{b}_j) = \sqrt{\mathrm{cov}(\hat{b}_{jj})}$	standard error of $\hat{b}_j$
$t_j = \hat{b}_j/SE(\hat{b}_j)$	test statistic for $\beta_j=0$
$p(t_j)$	probability associated with t_j (usually two tailed)

3. Case I: Ordinary Least Squares Definitions for Signed Permutation

Statistic	Description
$C = \{c_{ij}\} = X(X'X)^{-1}$	Nx(m+1) matrix of calculus or generalized inverse of X'
$\underset{\sim}{c}_j = \{c_{ij}\}$	The j^{th} column of C
$Q = \{q_{ii'}\} = C(C'C)^{-1}C' = X(X'X)^{-1}X'$	NxN idempotent matrix
$\hat{\underset{\sim}{b}} = \{b_j\} = C'\underset{\sim}{y}$	$(m+1)^{th}$ order vector of regression coefficients
$\hat{\underset{\sim}{y}} = \{\hat{y}_i\} = X\hat{\underset{\sim}{b}}$	N^{th} order vector of fitted values
$\underset{\sim}{e} = \{e_i\} = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$	N^{th} order vector of residuals
$\bar{e} = N^{-1}\Sigma_i e_i = 0$	Mean of residuals
$\mu_{2e} = \sigma^2 = N^{-1}\Sigma_i e_i^2$	Mean square or variance of residuals
$\underset{\sim}{y}_k = \{y_{ki}\} = \hat{\underset{\sim}{y}} + P_k\underset{\sim}{e}$	Constructed values of $\underset{\sim}{y}$ for the k^{th} signed permutation
$\hat{\underset{\sim}{b}}_k = \{b_{kj}\} = C'\underset{\sim}{y}_k$	Regression coefficients for the k^{th} signed permutation
$\hat{\underset{\sim}{y}}_k = \{\hat{y}_{ki}\} = X\underset{\sim}{b}_k$	Fitted values for the k^{th} signed permutation
$ave(\hat{\underset{\sim}{b}}_k) = \bar{\underset{\sim}{b}}' = K^{-1}\Sigma_k\hat{\underset{\sim}{b}}_k$	Average value of $\hat{b}_k$
$cov(\hat{\underset{\sim}{b}}_k) = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})'$	Covariance of $\hat{b}_k$
$d_k^2 = (\underset{\sim}{b}_k - \bar{\underset{\sim}{b}})(cov(\underset{\sim}{b}_k))^{-1}(\underset{\sim}{b}_k - \bar{\underset{\sim}{b}})$	Squared distance of $\hat{b}_k$ from $\bar{b}$

4. Case I: Ordinary Least Squares
Summary of K = 3840 Regression Equations

Regression Coefficient			Squared Distance	
Statistic	Intercept	Slope	Statistic	Distance
	b_{ko}	b_{k1}		d_k^2
$ave(b_{kj})$	2.0000	3.0000	$ave(d_k^2)$	2.0000
$var(b_{kj})$	29.4800	2.6800	$var(d_k^2)$	1.8058
$skew(b_{kj})$	0	0	--	--
$kurt(b_{kj})$	2.1500	2.1828	d^2	25.9328
z_j	.368	1.833	F*	12.966
$\tilde{p}(z_j)$	.35	.03	$\tilde{p}(F^*)$	$\cong 0$
$p(z_j)$	.3792	.0229	$p(F^*)$	0

NOTE: $z_j = b_j/\sqrt{var(b_{kj})}$, $\tilde{p}(z_j)$ from normal table, and $p(z_j)$ is compiled by dividing the number of b_{kj} with different signs by $K = 3840$.

d^2 is distance of b from origin, i.e. $\sigma^2 b'(C'C)^{-1}b$, F* is $d^2/(m + 1)$, $\tilde{p}(F^*)$ is from F table with $m + 1$ and ∞ d.f. $p(F^*)$ is computed by dividing the number of $d_k^2 > d^2$ by K .

5. **Case II: Weighting the Signed and Permuted Residuals Definitions**

Statistic	Description
$C = \{c_{ij}\} = WX(X'WX)^{-1}$	Nx(m+1) matrix of calculus or generalized inverse of X'
$c_j = \{c_{ij}\}$	The j^{th} column of c
$Q = \{q_{ii'}\} = C(C'C)^{-1}C'$	NxN idempotent matrix
$\hat{b} = \{b_j\} = C'y$	$(m+1)^{th}$ order vector of regression coefficients
$\hat{y} = \{\hat{y}_i\} = X\hat{b}$	N^{th} order vector of fitted values
$e = \{e_i\} = y - \hat{y}$	N^{th} order vector of residuals
$\bar{e} = N^{-1}\Sigma_i e_i$	Mean residual
$\sigma^2 = N^{-1}\Sigma_i(e_i - \bar{e})^2$	Variance of residuals
$\mu_{2e} = \sigma^2 + \bar{e}^2 = N^{-1}\Sigma_i e_i^2$	Mean square residual
$y_k = \{y_{ki}\} = \hat{y} + P_k e$	Constructed values of y for the k^{th} signed permutation
$\hat{b}_k = \{b_{kj}\} = C'y_k$	Regression coefficients for the k^{th} signed permutation
$\hat{y}_k = \{\hat{y}_{ki}\} = X\hat{b}_k$	Fitted values for the k^{th} signed permutation
$\text{ave}(\hat{b}_k) = \bar{b}' = K^{-1}\Sigma_k\hat{b}_k$	Average value of $\hat{b}_k$
$\text{cov}(\hat{b}_k) = K^{-1}\Sigma_k(\hat{b}_k - \bar{b})(\hat{b}_k - \bar{b})'$	Covariance of $\hat{b}_k$
$d_k^2 = (b_k - \bar{b})(\text{cov}(b_k))^{-1}(b_k - \bar{b})$	Squared distance of $\hat{b}_k$ from $\bar{b}$

6. Case II: Weighting the Resigned and Permuted Residuals
Summary of K = 3840 Regression Equations

Regression Coefficient			Squared Distance	
Statistic	Intercept	Slope	Statistic	Distance
	b_{ko}	b_{k1}		d_k^2
$\text{ave}(b_{kj})$	-3.3750	4.1250	$\text{ave}(d_k^2)$	2.0000
$\text{var}(b_{kj})$	57.4901	5.2080	$\text{var}(d_k^2)$	1.6500
$\text{skew}(b_{kj})$	--	--	--	--
$\text{kurt}(b_{kj})$	2.0165	2.0067	d^2	10.8958
z_j	.445	1.808	F*	5.448
$\tilde{p}(z_j)$	.33	.03	$\tilde{p}(F^*)$	$\cong 0$
$p(z_j)$	.3547	.0229	$p(F^*)$	0

NOTE: $z_j = b_j/\sqrt{\text{var}(b_{kj})}$, $\tilde{p}(z_j)$ from normal table, and $p(z_j)$ is compiled by dividing the number of b_{kj} with different signs by $K_k = 3840$.

d^2 is distance of b from origin, i.e. $\mu_{2e}b'(C'C)^{-1}b$, F* is $d^2/(m+1)$, $\tilde{p}(F^*)$ is from F table with $m+1$ and ∞ d.f. , $p(F^*)$ is computed by dividing the number of $d_k^2 > d^2$ by K_k .

7. Case III: Permuting the Weighted Residuals Definitions

Statistic	Description
$V = \{v_{ii}\}$	NxN matrix such that $V^2 = W$
$\underset{\sim}{y}^* = \{y_i^*\} = V\underset{\sim}{y}$	Weighted values of $\underset{\sim}{y}$
$X^* = \{x_{ij}^*\} = VX$	Weighted values of X
$C = \{c_{ij}\} = VX(X'WX)^{-1}$	Nx(m+1) matrix of calculus or generalized inverse of X'
$Q = \{q_{ii'}\} = C(C'C)^{-1}C'$	NxN idempotent matrix
$\underset{\sim}{\hat{b}} = \{\tilde{b}_j\} = C'V\underset{\sim}{y} = C'\underset{\sim}{y}^*$	$(m+1)^{th}$ order vector of regression coefficients
$\underset{\sim}{\hat{y}} = \{\underset{\sim}{\hat{y}}_i\} = X\underset{\sim}{\hat{b}}$	N^{th} order vector of fitted values
$\underset{\sim}{\hat{y}}^* = \{\hat{y}_i^*\} = X^*\underset{\sim}{\hat{b}} = V\underset{\sim}{\hat{y}}$	Weighted values of $\hat{y}$
$\underset{\sim}{e} = \{e_i\} = \underset{\sim}{y} - \underset{\sim}{\hat{y}}$	N^{th} order vector of residuals
$\underset{\sim}{e}^* = \{e_i^*\} = \underset{\sim}{y}^* - \underset{\sim}{\hat{y}}^* = V\underset{\sim}{e}$	Weighted values of residuals
$\bar{e}^* = N^{-1}\Sigma_i e_i^*$	Mean of weighted residuals
$\mu_{2e*} = N^{-1}\underset{\sim}{e}'W\underset{\sim}{e} = N^{-1}\Sigma_i(e_i^*)^2$	Mean square of weighted residuals
$\underset{\sim}{y}_k = \{y_{ki}\} = \underset{\sim}{\hat{y}} + V^{-1}P_kV\underset{\sim}{e}$	Fitted values for the k^{th} signed permutation

7. (Continued)

Statistic	Description
$\underset{\sim}{y}^*_k = \{y_{ki}\} = \hat{\underset{\sim}{y}}^* + P_k\underset{\sim}{e}^*$	Weighted values of $\underset{\sim}{y}_k$
$\underset{\sim}{b}_k = \{b_{kj}\} = C'V\underset{\sim}{y}_k = C'\underset{\sim}{y}_k$	Regression coefficients for the k^{th} signed permutation
$\hat{\underset{\sim}{y}}_k = \{\hat{y}_{ki}\} = X\hat{\underset{\sim}{b}}_k$	Fitted values for the k^{th} signed permutation
$\hat{\underset{\sim}{y}}^*_k = \hat{\underset{\sim}{y}}^*_{ki} = V\hat{\underset{\sim}{y}}_k$	Weighted values of $\hat{y}_k$
$\mathrm{ave}(\hat{\underset{\sim}{b}}_k) = \bar{\underset{\sim}{b}}' = K^{-1}\Sigma_k\hat{\underset{\sim}{b}}_k$	Average value of b_k
$\mathrm{cov}(\hat{\underset{\sim}{b}}_k) = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})'$	Covariance of b_k
$d_k^2 = N_k^{-1}(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})(\mathrm{cov}(\hat{\underset{\sim}{b}}_k))^{-1}(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})$	Squared distance of b_k from $\bar{b}$

8. Case III: Resigning and Permuting the Weighted Residuals
Summary of K = 3840 Regression Equations

Regression Coefficient			Squared Distance	
Statistic	Intercept	Slope	Statistic	Distance
	b_{ko}	b_{k1}		d_k^2
$ave(b_{kj})$	-3.3750	4.1250	$ave(d_k^2)$	2.0000
$var(b_{kj})$	30.7258	3.1219	$var(d_k^2)$	1.7422
$skew(b_{kj})$	0	0	--	--
$kurt(b_{kj})$	2.1728	2.1722	d^2	36.2613
z_j	.609	2.33	F*	18.13
$\tilde{p}(z_j)$	.27	.01	$\tilde{p}(F^*)$	$\cong$ 0
$p(z_j)$	.2862	0	p(F*)	0

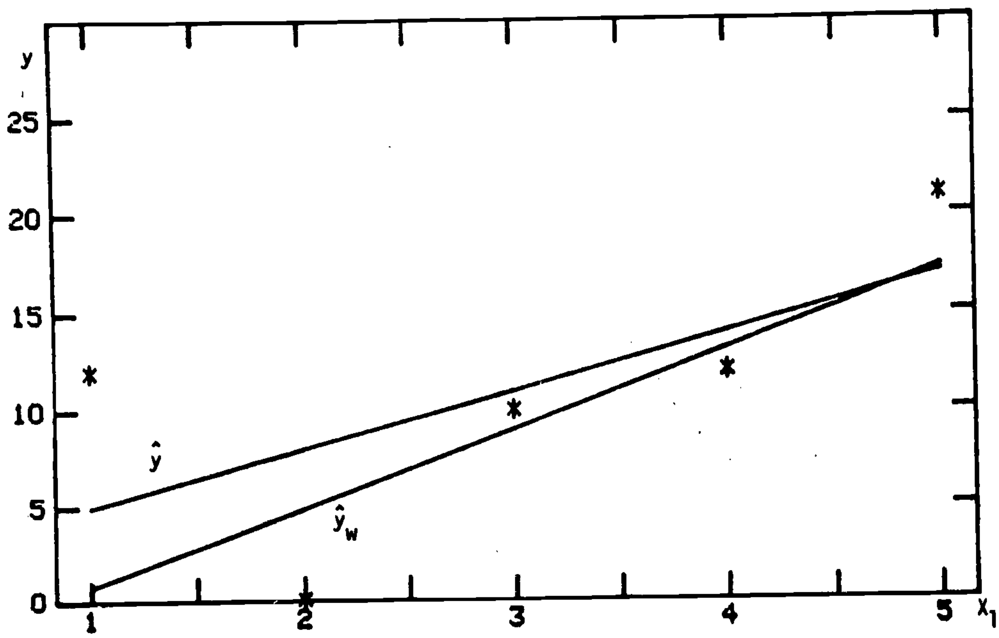
NOTE: $z_j = b_j/\sqrt{var(b_{kj})}$, $\tilde{p}(z_j)$ from normal table, and $p(z_j)$ is compiled by dividing the number of b_{kj} with different signs by K = 3840 .

d^2 is distance of b from origin, i.e. $\mu_{2e} b'(C'C)^{-1}b$, F* is $d^2/(m+1)$, $\tilde{p}(F^*)$ is from F table with m + 1 and ∞ d.f. , p(F*) is computed by dividing the number of $d_k^2 > d^2$ by K .

9. Summary of Signed Permutations and Moment Notation

Statistic	Description
	Signed Permutations
$K = 2^N N!$	Number of possible signed permutations
$k = 1,2,\ldots,K$	Index of signed permutations
$P_k = \{p_{kii'}\}$	NXN signed permutation matrix. Each row and column has exactly one nonzero element which may be either +1 or -1.
	Moment Notation
$\mu_{pz} = Z^{-1}\Sigma_j z_j^p$	The p^{th} (uncorrected) moment of variable z_j ($j = 1,2,\ldots,Z$)
$\beta_{1z} = \mu_{3z}^2/\mu_{2z}^3$	The (uncorrected) skewness of z
$\beta_{2z} = \mu_{4z}/\mu_{2z}^2$	The (uncorrected) kurtosis of z

A. Case I: Ordinary Least Squares
Numerical Example



Data

$$
\underset{\textstyle y}{\begin{bmatrix} 12 \\ 0 \\ 10 \\ 12 \\ 21 \end{bmatrix}}
\quad
\overset{\textstyle X}{\begin{array}{cc} x_0 & x_1 \end{array}}
\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}
\qquad
\sigma^2 \; [26.8] \qquad \underset{\sim}{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}
\qquad
\hat{\underset{\sim}{y}} = \begin{bmatrix} 5 \\ 8 \\ 11 \\ 14 \\ 17 \end{bmatrix}
\qquad
\underset{\sim}{e} = \begin{bmatrix} 7 \\ -8 \\ -1 \\ -2 \\ 4 \end{bmatrix}
\qquad
C = \begin{bmatrix} .8 & -.2 \\ .5 & -.1 \\ .2 & 0 \\ -.1 & .1 \\ -.4 & .2 \end{bmatrix}
$$

Note: The line $\hat{y}_w$ is the best fit weighted regression line and will be discussed in Section 3 .

B. Case I: Ordinary Least Squares Signed Permutations

Values of e_k and y_k

$$\underset{\sim}{e}_2 = \begin{bmatrix} -7 \\ -8 \\ -1 \\ -2 \\ -4 \end{bmatrix} \quad \underset{\sim}{y}_2 = \begin{bmatrix} -2 \\ 0 \\ 10 \\ 12 \\ 21 \end{bmatrix} \qquad \underset{\sim}{e}_3 = \begin{bmatrix} -8 \\ 7 \\ -1 \\ -2 \\ 4 \end{bmatrix} \quad \underset{\sim}{y}_3 = \begin{bmatrix} -3 \\ 15 \\ 10 \\ 12 \\ 21 \end{bmatrix} \qquad \underset{\sim}{e}_4 = \begin{bmatrix} 8 \\ 7 \\ 2 \\ -1 \\ -4 \end{bmatrix} \quad \underset{\sim}{y}_4 = \begin{bmatrix} 13 \\ 15 \\ 13 \\ 13 \\ 13 \end{bmatrix} \qquad \underset{\sim}{e}_5 = \begin{bmatrix} -8 \\ -4 \\ 1 \\ 2 \\ 7 \end{bmatrix} \quad \underset{\sim}{y}_5 = \begin{bmatrix} -3 \\ 4 \\ 12 \\ 16 \\ 24 \end{bmatrix}$$

Signed Permutation Matrices

$$P_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad P_4 = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad P_5 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Regression Coefficients

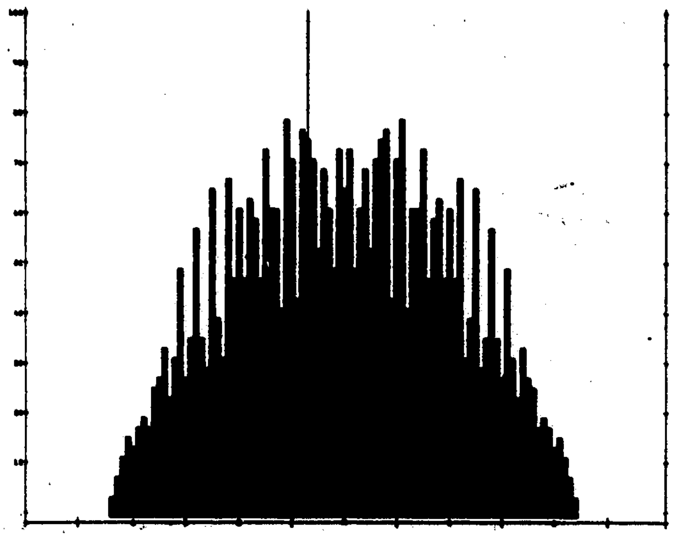
$$b_2 = \begin{bmatrix} -9.2 \\ 5.8 \end{bmatrix} \qquad b_3 = \begin{bmatrix} -2.5 \\ 4.5 \end{bmatrix} \qquad b_4 = \begin{bmatrix} 14 \\ -.2 \end{bmatrix} \qquad b_5 = \begin{bmatrix} -9.2 \\ 6.6 \end{bmatrix}$$

The minimum b_{kj} can be computed from the maximum by $\min b_{kj} = b_j - (\max b_{kj} - b_j)$ for example, the minimum intercept $= 2 - (14 - 2) = -10$.

CASE 1: UNWEIGHTED LEAST SQUARES

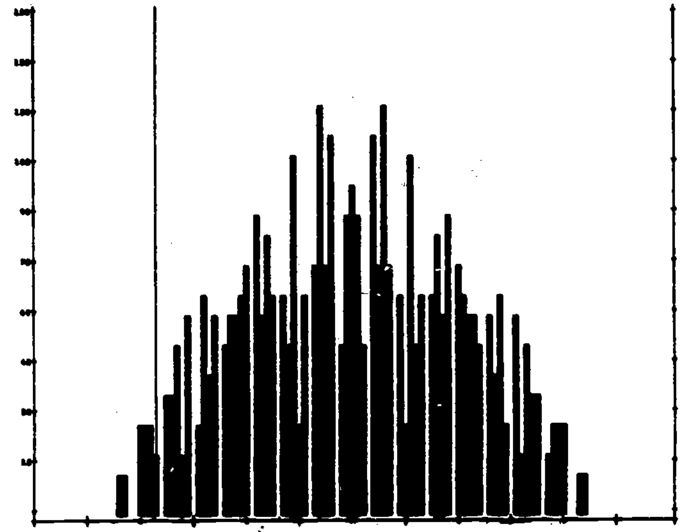
DISTRIBUTIONS OVER ALL SIGNED PERMUTATIONS

C1.



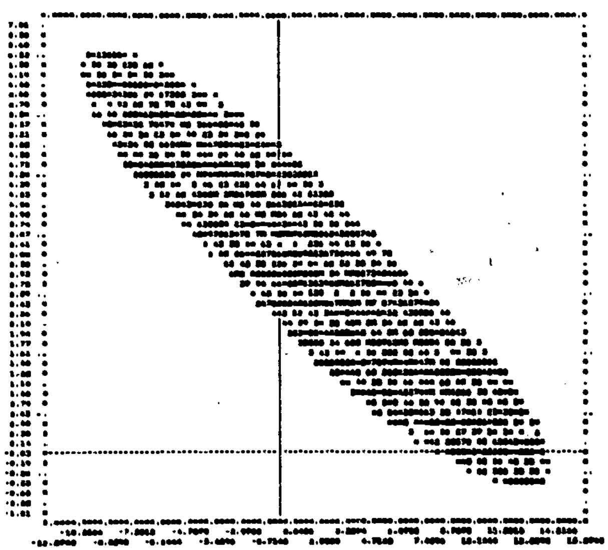
Distribution of b_{0k}

C2.



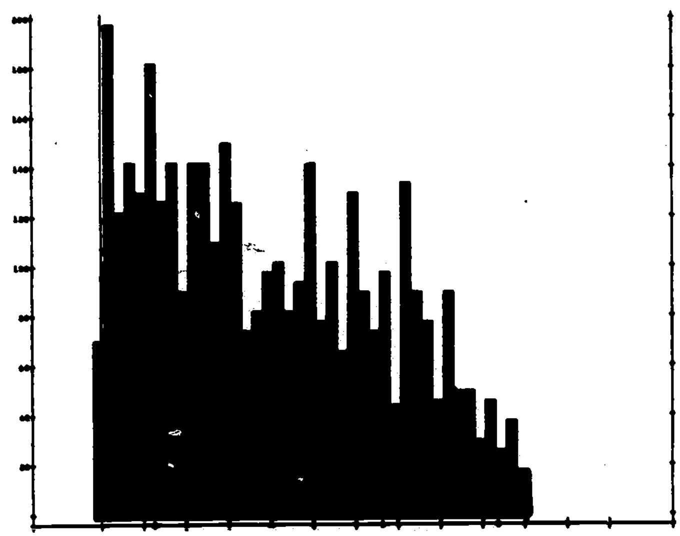
Distributution of b_{1k}

C3.



Bivariate Distribution of b_{0k} , b_{1k}

C4.



Distribution of d^2_k

D. Case II: Weighting the Permuted Residuals Regression Analysis

w_{ii}: $\begin{bmatrix} 1 \\ 4 \\ 9 \\ 4 \\ 1 \end{bmatrix}$

μ_{2e}: $\begin{bmatrix} 33.33 \end{bmatrix}$

$\hat{b}$: $\begin{bmatrix} -3.375 \\ 4.125 \end{bmatrix}$

$\hat{y}$: $\begin{bmatrix} .75 \\ 4.875 \\ 9.00 \\ 13.125 \\ 17.25 \end{bmatrix}$

e: $\begin{bmatrix} 11.25 \\ -4.875 \\ 1.00 \\ -1.125 \\ 3.75 \end{bmatrix}$

$C = WX(X'WX)^{-1}$: $\begin{bmatrix} .4276 & -.125 \\ .9605 & -.250 \\ .4737 & 0 \\ -.5395 & .250 \\ -.3224 & .125 \end{bmatrix}$

Signed Permutations

$e_2 = \begin{bmatrix} -11.25 \\ -4.88 \\ 1.00 \\ -1.13 \\ 3.75 \end{bmatrix}$ $y_2 = \begin{bmatrix} -10.50 \\ 0.00 \\ 10.00 \\ 12.00 \\ 21.00 \end{bmatrix}$ $e_3 = \begin{bmatrix} -4.88 \\ 11.25 \\ 1.00 \\ -1.13 \\ 3.75 \end{bmatrix}$ $y_3 = \begin{bmatrix} -4.13 \\ 16.13 \\ 10.00 \\ 12.00 \\ 21.00 \end{bmatrix}$ $e_4 = \begin{bmatrix} 1.13 \\ 11.25 \\ 3.75 \\ -4.88 \\ -1.00 \end{bmatrix}$ $y_4 = \begin{bmatrix} 1.88 \\ 16.13 \\ 12.75 \\ 8.25 \\ 16.25 \end{bmatrix}$ $e_5 = \begin{bmatrix} -3.75 \\ -11.25 \\ 1.00 \\ 4.88 \\ 1.13 \end{bmatrix}$ $y_5 = \begin{bmatrix} -3.00 \\ -6.38 \\ 10.00 \\ 18.00 \\ 18.38 \end{bmatrix}$

$P_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ $P_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ $P_4 = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$ $P_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}$

$b_2 = \begin{bmatrix} -13.00 \\ 6.94 \end{bmatrix}$ $b_3 = \begin{bmatrix} 5.22 \\ 2.11 \end{bmatrix}$ $b_4 = \begin{bmatrix} 12.65 \\ -0.17 \end{bmatrix}$ $b_5 = \begin{bmatrix} -18.30 \\ 8.77 \end{bmatrix}$

NOTE: The data are y and X shown in Figure A.

E. Case III: Permuting the Weighted Residuals Regression Analysis

$$\mu_{2e^*} \quad \begin{bmatrix} 49.95 \end{bmatrix} \qquad \hat{\underset{\sim}{b}} \quad \begin{bmatrix} -3.375 \\ 4.125 \end{bmatrix}$$

$$\hat{\underset{\sim}{y}}^* = \begin{bmatrix} .750 \\ 9.750 \\ 27.000 \\ 26.250 \\ 17.250 \end{bmatrix} \qquad \underset{\sim}{e}^* = \begin{bmatrix} 11.250 \\ -9.750 \\ 3.000 \\ -2.250 \\ 3.750 \end{bmatrix} \qquad C = VX(X'WX)^{-1} = \begin{bmatrix} .4276 & -.125 \\ .4803 & -.125 \\ .1579 & 0 \\ -.2697 & .125 \\ -.3224 & .125 \end{bmatrix}$$

Signed Permutations

Values of e_k and y_k

$$\underset{\sim 2}{e^*} = \begin{bmatrix} -11.25 \\ -9.75 \\ 3.00 \\ -2.25 \\ 3.75 \end{bmatrix} \quad \underset{\sim 2}{y^*} = \begin{bmatrix} -10.50 \\ 0.00 \\ 30.00 \\ 24.00 \\ 21.00 \end{bmatrix} \quad \underset{\sim 3}{e^*} = \begin{bmatrix} -9.75 \\ 11.25 \\ 3.00 \\ -2.25 \\ 3.75 \end{bmatrix} \quad \underset{\sim 3}{y^*} = \begin{bmatrix} -9.00 \\ 21.00 \\ 30.00 \\ 24.00 \\ 21.00 \end{bmatrix}$$

$$\underset{\sim 4}{e^*} = \begin{bmatrix} 9.75 \\ 11.25 \\ 2.25 \\ -3.00 \\ -3.75 \end{bmatrix} \quad \underset{\sim 4}{y^*} = \begin{bmatrix} 10.50 \\ 21.00 \\ 29.25 \\ 23.25 \\ 13.50 \end{bmatrix} \quad \underset{\sim 5}{e^*} = \begin{bmatrix} -11.25 \\ -9.75 \\ 2.25 \\ 3.75 \\ 3.00 \end{bmatrix} \quad \underset{\sim 5}{y^*} = \begin{bmatrix} -10.50 \\ 0.00 \\ 29.25 \\ 30.00 \\ 20.25 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad P_5 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\underset{\sim 2}{b} = \begin{bmatrix} 13.00 \\ 6.94 \end{bmatrix} \quad \underset{\sim 3}{b} = \begin{bmatrix} -2.27 \\ 4.13 \end{bmatrix} \quad \underset{\sim 4}{b} = \begin{bmatrix} 8.57 \\ 6.56 \end{bmatrix} \quad \underset{\sim 5}{b} = \begin{bmatrix} -14.49 \\ 7.59 \end{bmatrix}$$

NOTE: The data y and X are shown in Figure A and the weights are shown in Figure D.

APPENDIX

Lemmas

For the following lemmas, $\underset{\sim}{a}$ is an arbitrary column vector with elements a_i $(i = 1,2,\ldots,N)$, $\bar{a} = \Sigma a_i/N$, $\sigma_a^2 = \Sigma_i(a_i - \bar{a})^2/N$, $\mu_{2a} = \Sigma a_i^2/N = \sigma_a^2 + \bar{a}^2$, and $\mu_{4a} = \Sigma_i a_i^4/N$.

For all N^{th} order signed permutation matrices P_k, $k = 1,2,\ldots,K$ where $K = 2^N N!$, there is a signed permutation vector $\underset{\sim}{a}_k = P_k\underset{\sim}{a}$ with elements a_{ki}.

The subscript $i' = 1,2,\ldots,N$ but does not $= i$.

Lemma 1: $\Sigma_i \Sigma_{i'} a_i^2 a_{i'}^2 = N(N\mu_{2a}^2 - \mu_{4a})$

Proof:

$$\Sigma_i \Sigma_{i'} a_i^2 a_{i'}^2 = a_1^2 a_2^2 + a_1^2 a_3^2 + \ldots. + a_{N-1}^2 a_N^2$$

$$= a_1^2(\Sigma_i a_i^2 - a_1^2) + a_2^2(\Sigma_i a_i^2 - a_2^2) + \ldots + a_N^2(\Sigma_i a_i^2 - a_N^2)$$

$$= (\Sigma_i a_i^2)^2 - \Sigma_i a_i^4$$

$$= N[N\mu_{2a}^2 - \mu_{4a}] \quad .$$

Lemma 2: Given any set of elements $a_{ki_1}, a_{ki_2}, \ldots, a_{ki_m}$ $(i_1 \neq i_2 \neq \ldots \neq i_m)$ which are raised to integer powers $p_1, p_2, \ldots, p_m$ respectively, and at least one $P_j (j = 1,2,\ldots,m)$ is odd, then

$$K^{-1}\Sigma_k a_{ki_1}^{p_1} a_{ki_2}^{p_2} \ldots a_{ki_m}^{p_m} = 0 \quad .$$

Proof: If any power is odd, then the summation over all k signed permutations will contain each combination of other $a_{ki_j}^{p_j}$ an equal number of times with positive and negative signs.

Lemma 3: Let a_i be an element in $\underset{\sim}{a}$, then

a. $K^{-1}\Sigma_k a_{ki}^2 = \mu_{2a}$

b. $K^{-1}\Sigma_k a_{ki}^4 = \mu_{4a}$

c. $K^{-1}\Sigma_k a_{kj}^2 a_{ki'}^2 = (N - 1)^{-1}(N\mu_{2a}^2 - \mu_{4a})$

Proof:

a. $\Sigma_k a_{ki}^2$ contains the square of each element a_i of $\underset{\sim}{a}$ exactly $2^N(N - 1)!$ times, thus

$$K^{-1}\Sigma_k a_{ki}^2 = \frac{2^N(N-1)!}{2^N N!}\Sigma_i a_i^2 = \frac{\Sigma_i a_i^2}{N} = \mu_{2a} \quad .$$

b. $\Sigma_k a_{ki}^4$ contains the 4^{th} power of each element cf $\underset{\sim}{a}$ exactly $2^N(N - 1)!$ times, thus as 3a.

c. $\Sigma_k a_{ki}^2 a_{ki'}^2$ contains the product of squares of each distinct pair of elements in $\underset{\sim}{a}$ exactly $2^N(N - 2)!$ times, thus

$$K^{-1}\Sigma_k a_{ki}^2 a_{ki'}^2 = \frac{2^N(N-2)!}{2^N N!}[\Sigma_i \Sigma_{i'} a_i^2 a_{i'}^2] \quad .$$

Using Lemma 1 on the term in brackets,

$$= \frac{1}{N(N-1)} [N(N\mu_{2a}^2 - \mu_{4a})]$$

$$= (N-1)^{-1}(N\mu_{2a}^2 - \mu_{4a}) \quad .$$

Lemma 4: Given a column vector $\underset{\sim}{c}$ with elements c_i $(i = 1,2,\ldots,N)$ with (uncorrected) moments $\mu_{1c} = \Sigma_i c_i/N$, $\mu_{2c} = \Sigma_i c_i^2/N$, and $\mu_{4c} = \Sigma_i c_i^4/N$, and p an odd positive integer, then

a. $K^{-1}\Sigma_k(\underset{\sim}{c}'\underset{\sim}{a}_k)^p = 0$

b. $K^{-1}\Sigma_k(\underset{\sim}{c}'\underset{\sim}{a}_k)^2 = N\mu_{2a}\mu_{2c}$

c. $K^{-1}\Sigma_k \underset{\sim}{c}'\underset{\sim}{a}_k)^4 = \dfrac{\mu_{4a}\mu_{4c}}{N} + \dfrac{3N}{N-1}(N\mu_{2a}^2 - \mu_{4a})(N\mu_{2c}^2 - \mu_{4c})$

Proof:

a. Each term of $(c_1a_{k1} + c_2a_{k2} + \ldots + c_Na_{kN})^p$ will have at least one odd power, thus each term vanishes by Lemma 2.

b. Using Lemma 3a, $K^{-1}\Sigma_k(c_1^2a_{k1}^2 + c_2^2a_{k2}^2 + \ldots + c_N^2a_{kN}^2 + \text{odd powers})$
$= c_1^2\mu_{2a}' + c_2^2\mu_{2a}' + \ldots. + c_N^2\mu_{2a}' = N\mu_{2a}\mu_{2c}$.

c. $K^{-1}\Sigma_k(\underset{\sim}{c}\underset{\sim}{a}_k)^4$ has terms $c_i^4a_{ki}^4, 3c_ic_{i'}, a_{ki}^2a_{ki'}^2$, and odd powers. Using Lemmas 3b and 3c on the terms in brackets,

$$K^{-1}\Sigma_k(\underset{\sim}{c}\underset{\sim}{a}_k)^4 = c_1^4[K^{-1}\Sigma_k a_{k1}^4] + c_2^4[K^{-1}\Sigma_k a_{k2}^4] + \ldots$$

$$+ 3c_1^2c_2^2[K^{-1}\Sigma_k a_{k1}^2a_{k2}^2] + \ldots. + \text{odd powers}$$

$$= \mu_{4a}\Sigma_i c_i^4 + \frac{3}{N-1}(N\mu_{2a}^2 - \mu_{4a})\Sigma_i\Sigma_{i'}c_i^2c_{i'}^2$$

$$= N\mu_{4a}\mu_{4c} + \frac{3N}{N-1}(N\mu_{2a}^2 - \mu_{4a})(N\mu_{2c}^2 - \mu_{4c}) \quad .$$

Lemma 5. For the vectors $\underset{\sim}{a}_k$,

a. $K^{-1}\Sigma_k \underset{\sim}{a}_k = K^{-1}\Sigma_k P_k \underset{\sim}{a} = 0$

b. $K^{-1}\Sigma_k \underset{\sim}{a}_k \underset{\sim}{a}'_k = K^{-1}\Sigma_k P_k \underset{\sim}{a}\underset{\sim}{a}' P'_k = \mu'_{2a} I = (\sigma_a^2 + \bar{a}^2) I$

where I is an N^{th} order identity matrix.

Proof:

a. Each element of $\underset{\sim}{a}$ is contained in the sum an equal number of times with positive and negative signs.

b. The diagonal elements of $K^{-1}\Sigma_k \underset{\sim}{a}_k \underset{\sim}{a}'_k$ are $K^{-1}\Sigma_k a_{ki}^2 = \mu_{2a}$ and the off diagonals are odd powered. Also, $\mu'_{2a} = \sigma_a^2 + \bar{a}^2$.

Lemma 6: Given an N x N matrix Q of rank $r \leq N$ then

$$\text{a)} \quad K^{-1} \Sigma_k P_k Q P_k' = \frac{Tr(Q)}{N} I$$

and if Q is idempotent (i.e. QQ = Q) then

$$\text{b)} \quad K^{-1} \Sigma_k P_k Q P_k' = \frac{r}{N} I$$

where I is an Nth order identity matrix.

Proof: Since each off-diagonal element of Q, $q_{ii'}$, say, is matched with $-q_{ii'}$ in all off-diagonal summations, then $K^{-1}\Sigma_k P_k Q P_k'$ is at least diagonal. Each diagonal element of $K^{-1}\Sigma_k P_k Q P_k'$ consists of the sum of the diagonal element of Q exactly $2^N(N-1)!$ times, thus $K^{-1}\Sigma_k P_k Q P_k' = N^{-1}Tr(Q)I$. If $QQ = Q$, then $Tr(Q) = r$ and $K^{-1} \Sigma_k P_k Q P_k' = N^{-1} r I$.

Theorem 1 (Case I: Ordinary Least Squares)

Given the definitions in Tables 1, 3, and 9, then

a. $\text{ave}(\hat{\underset{\sim}{b}}_k) = \hat{\underset{\sim}{b}}$

b. $\text{cov}(\hat{\underset{\sim}{b}}_k) = \mu_{2e}C'C = \sigma^2(X'X)^{-1}$

c. $\text{skew}(\hat{b}_{kj}) = 0$

d. $$\text{kurt}(\hat{b}_{kj}) = \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right)$$

e. $\text{ave}(d_k^2) = m + 1$

f. $$\text{var}(d_k^2) = \left(\frac{1 - \beta_{2e}}{N - 1}\right)(m + 1)^2 + 2\left(\frac{N - \beta_{2e}}{N - 1}\right)(m + 1)$$
$$+ \Sigma_i q_{ii}^2\left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1}\right)$$

g. $\text{ave}(\hat{\underset{\sim}{y}}_k) = \hat{\underset{\sim}{y}}$

h. $\text{cov}(\hat{\underset{\sim}{y}}_k) = \mu_{2e}XC'CX' = \sigma^2X(X'X)^{-1}X'$

Note that all moments are central since $\bar{e} = 0$ and $\Sigma_i c_{ij} = 0$ for all j except j = 0, the intercept.

Proof: Note that $\hat{\underset{\sim}{b}}_k - \hat{\underset{\sim}{b}} = C'P_k\underset{\sim}{e}$ and $\hat{y}_k - \hat{y} = XC'P_k\underset{\sim}{e}$,

a. Using Lemma 5a on the term in brackets,

$$\text{ave}(\hat{\underset{\sim}{b}}_k) = K^{-1}\hat{\underset{\sim}{b}}_k = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}} + C'P_k\underset{\sim}{e})$$
$$= \hat{\underset{\sim}{b}} + C'[K^{-1}\Sigma_kP_k\underset{\sim}{e}] = \hat{\underset{\sim}{b}}$$

b. Using Lemma 5b and the fact that $\bar{\underset{\sim}{b}} = \hat{\underset{\sim}{b}}$,

$$\mathrm{cov}(\hat{\underset{\sim}{b}}_k) = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})'$$

$$= K^{-1}\Sigma_k(C'P_k\underset{\sim}{e})(C'P_k\underset{\sim}{e})'$$

$$= C'[K^{-1}\Sigma_k P_k\underset{\sim}{e}\underset{\sim}{e}'P_k']C = C'(\mu_{2e}I)C'$$

$$= \mu_{2e}C'C$$

since $\mu_{2e} = \sigma^2$ and $C'C = (X'X)^{-1}$, then

$$\mathrm{cov}(\hat{\underset{\sim}{b}}_k) = \sigma^2(X'X)^{-1} \quad .$$

c. Noting that $\underset{\sim}{c}_j$ is the j^{th} column of C and using Lemma 4a,

$$\mathrm{skew}(\hat{b}_{kj}) = \beta_{1j} = (K^{-1}\Sigma_k(\hat{b}_{kj} - \hat{b}_j)^3)^2/(K^{-1}\Sigma_k(\hat{b}_{kj} - \hat{b}_j)^2)^3$$

$$= [K^{-1}\Sigma_k(\underset{\sim}{C}_j'P_k\underset{\sim}{e})^3]^2/(K^{-1}(\Sigma_k\underset{\sim}{C}_j'P_k\underset{\sim}{e})^2)^3$$

$$= 0 \quad .$$

d. Using Lemmas 4b and 4c,

$$\text{kurt}(\hat{b}_{kj}) = \beta_{2j} = K^{-1}\Sigma_k(\hat{b}_{kj} - \hat{b}_j)^4/(K^{-1}\Sigma_k(\hat{b}_{kj} - \hat{b}_j)^2)^2$$

$$= [K^{-1}\Sigma_k(\underset{\sim}{c}_j'P_k\underset{\sim}{e})^4]/[K^{-1}\Sigma_k(\underset{\sim}{c}_j'P_k\underset{\sim}{e})^2]^2$$

$$= \frac{N\mu_{4e}\mu_{4c} + \frac{3N}{N-1}(N\mu_{2e}^2 - \mu_{4e})(N\mu_{2c}^2 - \mu_{4cj})}{(N\mu_{2c}\mu_{2c})^2}$$

$$= \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right) \quad .$$

e. Using Lemma 6b,

$$\text{ave}(d_k^2) = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})'(\text{cov}(\hat{\underset{\sim}{b}}_k))^{-1}(\hat{\underset{\sim}{b}}_k - \bar{\underset{\sim}{b}})$$

$$= K^{-1}\Sigma_k(C'P_k\underset{\sim}{e})'(\mu_{2e}C'C)^{-1}(C'P_k\underset{\sim}{e})$$

$$= \mu_{2e}^{-1}\underset{\sim}{c}'(K^{-1}\Sigma_k P_k'C(C'C)^{-1}C'P_k)\underset{\sim}{e}$$

$$= \mu_{2e}^{-1}\underset{\sim}{e}'[K^{-1}\Sigma_k P_k'QP_k]\underset{\sim}{e}$$

$$= \mu_{2e}^{-1}\underset{\sim}{e}'\left(\frac{m+1}{N}I\right)\underset{\sim}{e} = \frac{m+1}{\mu_{2e}}\frac{\underset{\sim}{e}'\underset{\sim}{e}}{N}$$

$$= m + 1 \quad .$$

f. Since $d_k^2 = \mu_{2e}^{-1} e' P_k' Q P_k e$,

$$\mathrm{var}(d_k^2) = K^{-1}\Sigma_k d_k^4 - (\mathrm{ave}(d_k^2))^2$$

$$= K^{-1}\Sigma_k (\mu_{2e}^{-1} e' P_k Q P_k e)^2 - (m + 1)^2$$

$$= \mu_{2e}^{-2} K^{-1} \Sigma_k (q_{11} e_{k1}^2 + q_{12} e_{k1} e_{k2} + \cdots$$

$$+ q_{NN} e_{kN}^2)^2 - (m + 1)^2 \quad .$$

After squaring the term in the first parentheses and rearrangement,

$$\mathrm{var}(d_k^2) = \mu_{2e}^{-2} K^{-1} (\Sigma_i q_{ii}^2 \Sigma_k e_{ki}^4 + 2\Sigma_i \Sigma_{i'} q_{ii'}^2 \Sigma_k e_{ki}^2 e_{ki'}^2$$

$$+ \Sigma_i \Sigma_{i'} q_{ii} q_{i'i'} \Sigma_k e_{ki}^2 e_{ki'}^2 + \text{odd powers})$$

$$- (m + 1)^2 \quad .$$

By Lemma 3b,

$$\mu_{2e}^{-2} \Sigma_i q_{ii}^2 [K^{-1} \Sigma_k e_{ki}^4] = \frac{\mu_{4e}}{\mu_{2e}^2} \Sigma_i q_{ii}^2 \approx \beta_{2e} \Sigma_i q_{ii}^2 \quad .$$

By Lemma 3c and the fact that $\Sigma_i\Sigma_{i'}q^2_{ii'} = (m + 1) - \Sigma_i q^2_{ii}$,

$$2\mu_{2e}^{-2}K^{-1}\Sigma_i\Sigma_{i'}q^2_{ii'}\Sigma_k e^2_{ki}e^2_{ki'} = 2\mu_{2e}^{-2}\Sigma_i\Sigma_{i'}q^2_{ii'}[K^{-1}\Sigma_k e^2_{ki}e^2_{ki'}]$$

$$= \frac{2}{\mu^2_{2e}(N - 1)}(N\mu^2_{2e} - \mu_{4e})\Sigma_i\Sigma_{i'}q^2_{ii'}$$

$$= \frac{2}{N - 1}(N - \beta_{2e})((m + 1) - \Sigma_i q^2_{ii}) \quad .$$

By Lemma 3c and the fact that $\Sigma_i\Sigma_{i'}q_{ii}q_{i'i'} = (m + 1)^2 - \Sigma_i q^2_{ii}$,

$$\mu_{2e}^{-2}K^{-1}\Sigma_i\Sigma_{i'}q_{ii}q_{i'i'}\Sigma_k e^2_{ki}e^2_{ki'} = \mu_{2e}^{-2}\Sigma_i\Sigma_{i'}q_{ii}q_{i'i'}[K^{-1}\Sigma_k e^2_{ki}e^2_{ki'}]$$

$$= \frac{1}{\mu^2_{2e}(N - 1)}(N\mu^2_{2e} - \mu_{4e})\Sigma_i\Sigma_{i'}q_{ii}q_{i'}q_{i'}$$

$$= \frac{1}{N - 1}(N - \beta_{2e})((m + 1)^2 - \Sigma_i q^2_{ii}) \quad .$$

By Lemma 2 all odd powers vanish, thus

$$\mathrm{var}(d^2_k) = \beta_{2e}\Sigma_i q^2_{ii} + \frac{2}{N - 1}(N - \beta_{2e})(m + 1 - \Sigma_i q^2_{ii})$$

$$+ \frac{1}{N - 1}(N - \beta_{2e})((m + 1)^2 - \Sigma_i q^2_{ii}) - (m + 1)^2 \quad .$$

Rearrange in terms in powers of $m + 1$,

$$\mathrm{var}(d_k^2) = \frac{1 - \beta_{2e}}{N - 1}(m + 1)^2 + \frac{2(N - \beta_{2e})}{N - 1}(m + 1)$$

$$+ \Sigma_i q_{ii}^2 \left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1} \right) \quad .$$

g. Using Lemma 5a,

$$\mathrm{ave}(\hat{\underset{\sim}{y}}_k) = K^{-1}\Sigma_k \hat{\underset{\sim}{y}}_k = K^{-1}\Sigma_k(\hat{\underset{\sim}{y}} + XC'P_k\underset{\sim}{e})$$

$$= \hat{\underset{\sim}{y}} + XC'[K^{-1}\Sigma_k P_k \underset{\sim}{e}]$$

$$= \hat{\underset{\sim}{y}} \quad .$$

h. Using Lemma 5b,

$$\mathrm{cov}(\hat{\underset{\sim}{y}}_k) = K^{-1}\Sigma_k(\hat{\underset{\sim}{y}}_k - \hat{\underset{\sim}{y}})(\hat{\underset{\sim}{y}}_k - \hat{\underset{\sim}{y}})'$$

$$= K^{-1}\Sigma_k(XC'P_k\underset{\sim}{e})(XC'P_k\underset{\sim}{e})'$$

$$= XC'[K^{-1}\Sigma_k P_k \underset{\sim}{e}\underset{\sim}{e}'P_k']CX'$$

$$= \mu_{2e}XC'CX'$$

and thus

$$\mathrm{cov}(\hat{\underset{\sim}{y}}_n) = \sigma^2 X(X'X)^{-1}X' \quad .$$

Corollary 1:

Given a subset $\hat{\underset{\sim}{b}}_s$ of $\hat{\underset{\sim}{b}}$, the corresponding subset $\hat{\underset{\sim}{b}}_{ks}$ of $\hat{\underset{\sim}{b}}_k$, and the subset $\mathrm{cov}(\hat{\underset{\sim}{b}}_{ks})$ of $\mathrm{cov}(\hat{\underset{\sim}{b}}_k)$, $d^2_{ks} = (\hat{\underset{\sim}{b}}_{sk} - \hat{\underset{\sim}{b}}_s)'(\mathrm{Cov}(\hat{\underset{\sim}{b}}_{ks}))^{-1}(\hat{\underset{\sim}{b}}_{sk} - \hat{\underset{\sim}{b}}_s)$, then the

$$\mathrm{ave}(d^2_{ks}) = K^{-1}\Sigma_k(\hat{\underset{\sim}{b}}_{ks} - \hat{\underset{\sim}{b}}_s)'(\mathrm{cov}(\hat{\underset{\sim}{b}}_{ks}))^{-1}(\hat{\underset{\sim}{b}}_{ks} - \hat{\underset{\sim}{b}}_s) = m_s$$

and

$$\mathrm{var}(d^2_{ks}) = \left(\frac{1 - \beta_{ks}}{N - 1}\right)m_s^2 + 2\left(\frac{N - \beta_{2e}}{N - 1}\right)m_s$$

$$+ \Sigma_i q^2_{sii}\left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1}\right)$$

where m_s is the number of elements in $\hat{\underset{\sim}{b}}_s$ and the q_{sii} are defined below.

Proof:

Let X be partitioned into $(X_s, X_{\bar{s}})$ where X_s in the N x m_s matrix consisting of the columns of X corresponding to $\hat{\underset{\sim}{b}}_s$ and $X_{\bar{s}}$ is a N x $m_{\bar{s}}$ ($m_{\bar{s}} = m + 1 - m_s$) matrix containing the remaining columns. Also, let

$$\tilde{X}_s = X_s - X_{\bar{s}}(X_{\bar{s}}'X_{\bar{s}})^{-1}X_{\bar{s}}'Xs$$

and

$$\tilde{y}_s = y - X_s(X_s'X_s)^{-1}X_s'y \quad .$$

These values may be substituted for X and $\underset{\sim}{y}$ in Table 4 without affecting the values of $\tilde{y}$ and $\underset{\sim}{e}$. Thus, C becomes

$$C_s = \tilde{X}_s(\tilde{X}_s'\tilde{X}_s)'$$

and

$$C_s'\tilde{y}_s = \hat{b}_s \quad .$$

With these substitutions in Table 4, Theorem 1e and 1f follow. The matrix Q becomes

$$Q_s = C_s(C_s'C_s)^{-1}C_s' = \tilde{X}_s(\tilde{X}_s'\tilde{X}_s)^{-1}\tilde{X}_s'$$

where Q_s has elements $q_{sii'}$ $(i,i' = 1,2,\ldots,N)$. The value of m + 1 becomes m_s since m_s is the rank of Q_s

Note: In most statistical analyses the variance of the intercept b_o is not of interest; thus $\tilde{X}_s$ and $\tilde{y}_s$ are X and $\underset{\sim}{y}$ centered about their respective means and the ave $d_k^2 = m$.

Theorem 2: (Case II: Permuting the Weighted Residuals)

Given the definitions in Tables 1, 5, and 9, then

a. $\text{ave}(\hat{b}_k) = \hat{b}$

b. $\text{cov}(\hat{b}_k) = \mu_{2e}C'C = (\sigma_e^2 + \bar{e}^2)WX(X'WX)^{-1}X'W^2X(X'WX)^{-1}X'W$

c. $\text{skew}(\hat{b}_{kj}) = 0$

d. $$\text{kurt}(\hat{b}_{kj}) = \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right)$$

e. $\text{ave}(d_k^2) = m + 1$

f. $$\text{var}(d_k^2) = \left(\frac{1 - \beta_{2e}}{N - 1}\right)(m + 1)^2 + 2\left(\frac{N - \beta_{2e}}{N}\right)(m + 1)$$
$$+ \Sigma_i q_{ii}^2\left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1}\right)$$

g. $\text{ave}(\hat{y}_k) = \hat{y}$

h. $\text{cov}(\hat{y}_k) = \mu_{2e}XC'CX' = (\sigma_e^2 + \bar{e}^2)X(X'WX)^{-1}X'W^2X(X'WX)^{-1}X'$

Note that the moments are not in general central except for

μ_{pc_j} $(p > 0, j > 0)$.

Proof: With the substitution of definitions from Table 5 for Table 4, the proof follows the same steps as Theorem 1 except that

$$\mu_{2e} = \sigma_e^2 + \bar{e}^2$$

and

$$C'C = WX(X'WX)^{-1}X'W^2X(X'WX)^{-1}X'W \quad .$$

Theorem 3: (Case III: Permuting the Weighted Residuals)

Given the definitions in Table 1, 7, and 9, then

a. $\text{ave}(\hat{\underset{\sim}{b}}_k) = \hat{\underset{\sim}{b}}$

b. $\text{cov}(\hat{\underset{\sim}{b}}_k) = \mu_{2e}C'C - \left(\frac{\underset{\sim}{e}'W\underset{\sim}{e}}{N} \right)(X'WX)^{-1}$

c. $\text{skew}(\hat{b}_{kj}) = 0$

d. $\text{kurt}(\hat{b}_{kj}) = \frac{\beta_{2e}\beta_{2c_j}}{N} + \frac{3N}{N-1}\left(1 - \frac{\beta_{2e}}{N}\right)\left(1 - \frac{\beta_{2c_j}}{N}\right)$

e. $\text{ave}(d_k^2) = m + 1$

f. $\text{var}(d_k^2) = \left(\frac{1 - \beta_{2e}}{N - 1} \right)(m + 1)^2 + 2\left(\frac{N - \beta_{2e}}{N - 1} \right)(m + 1)$

$$+ \Sigma_i q_{ii}^2 \left(\frac{\beta_{2e}(N + 2)}{N - 1} - \frac{3N}{N - 1} \right)$$

g. $\text{ave}(\hat{\underset{\sim}{y}}_k) = \hat{\underset{\sim}{y}}$

h. $\text{cov}(\hat{\underset{\sim}{y}}_k) = \mu_{2e}XC'CX' = \left(\frac{\underset{\sim}{e}'w\underset{\sim}{e}}{N} \right)X(X'WX)^{-1}X'$

Note that the moments are not, in general, central.

Proof: With the substitution of definitions from Table 6 for Table 4 and substitution of $\underset{\sim}{y}^*$, X^*, $\hat{\underset{\sim}{y}}^*$, $\underset{\sim}{e}^*$, $\bar{e}^*$, μ_{2e^*}, $\underset{\sim}{y}_k^*$, and $\hat{\underset{\sim}{y}}_k^*$ for their unstarred equivalents, then the proof follows the same steps as Theorem 1 except that

$$\mu_{2e^*} = \frac{\underset{\sim}{e}'w\underset{\sim}{e}}{N}$$

and

$$C'C = (X'WX)^{-1} \quad .$$

REFERENCES

Beaton, Albert E., and Tukey, John W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," Technometrics, 16, 2, 147-185.

Daniel, Cuthbert, and Wood, Frederick S. (1971), "Fitting Equations to Data," Interscience, New York: John Wiley.

Draper, Norman R., and Smith, H. (1966), Applied Regression Analysis, New York: John Wiley.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," The Annals of Statistics, 7, 1, 1-26.

Fisher, R. A. (1926), Applications of "Student's Distribution" Metron, 5, 4, 99-104.

——————— (1930), Statistical Methods for Research Workers, London: Oliver and Boyd, Third Edition.

Freedman, David, and Lane, David (1978), "Significance Testing in a Nonstochastic Setting," Technical Report No. 317.

Graybill, Franklin A. (1961), An Introduction to Linear Statistical Models, New York: McGraw-Hill.

Lustig, Roger L. (1979), "Regression Analysis with Signed Permutations," Senior Thesis, Princeton University.

Mosteller, Frederick, and Tukey, John W. (1977), Data Analysis and Regression, Reading, Massachusetts: Addison Wesley.

Pitman, E. H. G. (1937), "Significance Tests Which May Be Applied to Samples From Any Populations," Journal of the Royal Statistical Society, Series B, 4, 119-130.

Searle, S. R. (1971), Linear Models, New York; John Wiley and Sons.